What you *see* is not what you *hear*

Word count: 4867

Furkan Dikmen
Université Côte d'Azur, CNRS, UMR7320:BCL
25 avenue François Mitterrand, 06300 Nice, France
furkan.dikmen@etu.univ-cotedazur.fr

Lena Baunaz
Université Côte d'Azur, CNRS, UMR7320:BCL
25 avenue François Mitterrand, 06300 Nice, France
lena.baunaz@univ-cotedazur.fr

Abstract This squib addresses an asymmetry between French perception verbs *voir* 'see' and *entendre* 'hear'. While both verbs license ECM configurations with overt embedded subjects, only *entendre* allows infinitival complements without an overt subject. We propose that this asymmetry is rooted within the semantic types of objects that these verbs combine with, correlating with the size of the structure they embed.

Keywords: perception, French, semantic objects, syntactic size

1 Puzzle

It is well-known that perception verbs like *hear* and *see* in at least Romance permit ECM complements when the embedded predicate has an overt subject, as shown in (1).

(1) a. J'ai entendu [Nicolas ronfler].
I.have heard Nicolas snore.INF
'I heard Nicolas snoring.'
b. J'ai vu [Nicolas marcher].
I.have seen Nicolas walk.INF
'I saw Nicolas walking.'

Our novel observation is that while both verbs in French license ECM configurations with overt embedded subjects, only *entendre* allows for infinitival complements without an overt subject (see also Dikmen et al. 2025).[1]

[1] Let us note in passing that the embedded material in (2a) is not a syntactically nominal element but a verbal one. First, they cannot co-occur with a determiner (neither definite nor indefinite), (i) (compare with (ib)). Second, adjectives cannot act as modifiers to the embedded form while they are compatible with its nominal form as shown in (ii). Furthermore, they cannot occur in the restrictor of a quantifier as opposed to its nominal form as shown in (iii).

(i) a. *J'ai entendu le/la/un/une ronfler.
I.have heard the.M.SG/the.FEM.SG/a.M.SG/the.M.SG snore.INF
Int: 'I heard the/a snoring.'
b. J'ai entendu le/un ronflement.
I.have heard the/a snoring
'I heard a snoring.'

(2) a. J'ai entendu [ronfler].
I.have heard snore.INF
≈'I heard someone snoring.'
Lit: I heard snore.
b. *J'ai vu [marcher].
I.have seen walk.INF
Int(ended): ≈'I saw someone walking.'

In particular, (2a) roughly means that the speaker auditorily perceived a snoring event, which we approximate in English as 'the speaker heard someone snoring/the speaker heard some snoring'. Importantly, the corresponding interpretation is not available with *voir*. In other words, (2b) is ungrammatical and thus cannot mean that the speaker saw someone walking or some walking. This contrast is unexpected because if the interpretation of (2a) was derived by positing an arbitrary covert subject such as pro_{arb}, one might expect *voir* to behave similarly, since both verbs otherwise pattern alike in ECM contexts.

2 Size of the embedded clause

In order to identify the syntax of the contrasting examples in (2), we will present additional data points. First, we observe that neither *entendre* nor *voir* can embed an unaccusative verbal head directly, without its object being present, as shown in (3a)-(3b).

(3) a. *Jean a entendu/vu tomber.
Jean has heard/seen fall.INF
'John heard/saw falling.'
b. Jean a entendu/vu (le vase) tomber (le vase).
Jean has heard/seen the vase fall.INF the vase
'John heard/saw the vase fall.'

Second, we note that when *entendre* embeds a predicate without its external argument, the embedded material cannot be modified with agent oriented adverbs or purpose (control) clauses, as shown in (4c), suggesting that these constructions lack a syntactically projected agent. This contrasts sharply with sentences like (4a) and (4b), where *voir* and *entendre* respectively embeds a logical subject, which can license purpose clauses and agent oriented adverbs. This means that when *entendre* embeds infinitivals without an overt subject as in (2a), there is no structure to host an implicit external argument either. We will take this to indicate that such structures lack vP projections.[2]

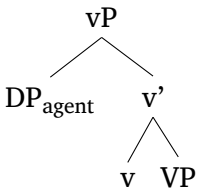
(ii) a. J'ai entendu (?doucement/*doux) ronfler (doucement/*doux).
I.have heard (softly/soft) snoring (softly/soft)
Int: 'I heard soft snoring.'
b. J'ai entendu un doux ronflement.
I.have heard a soft snoring
'I heard a soft snoring.'

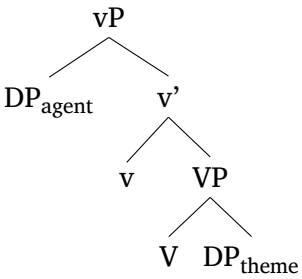
(iii) a. J'ai entendu chaque ronflement de Jean.
I.have heard every snoring of Jean
'I heard every snoring of Jean.'
b. *J'ai entendu chaque ronfler de Jean.
I.have heard every snoring of Jean
Int: 'I heard every snoring of jean.'

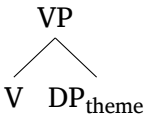
[2] Unlike other Romance languages like Spanish and Italian, French is not a pro drop language. This might be why French disallows implicit pro subjects in these constructions. What is important for us here is the

(4) a. Hier, j'ai vu Nicolas gesticuler pour alerter la police/délibérément.
yesterday I.have seen Nicolas wave.INF to alert the police/deliberately
'Yesterday I saw Nicolas waving to alert the police/deliberately.'
b. Jean a entendu Marie crier pour déranger les voisins/délibérément.
Jean has heard Marie scream.INF to disturb the neighbors/deliberately
'John heard Mary shouting to disturb the neighbors / on purpose.'
c. *Jean a entendu crier pour déranger les voisins/délibérément.
Jean has heard scream.INF to disturb the neighbors/deliberately
Int: 'John heard someone shouting to disturb the neighbors / on purpose.'

We adopt the standard assumptions about argument structure that the single argument of an unergative verb is merged in in [Spec, vP], where vP is the projection immediately above VP (5a) (Kratzer 1996 a.o.). Transitives share the unergative structure but include an additional theme argument combining with the head V, (5b). The sole argument of unaccusatives is projected as the complement of the head V (5c) (Perlmutter 1978).

(5) a. 

b. 

c. 

The data in (3) and (4) combined with the argument structures in (5) suggest that *entendre* can embed structures that lack logical subjects, such as VPs, whereas *voir* cannot. This has two consequences: (i) when *voir* embeds unaccusatives, there must be higher projections on top of their VP level, as will be clarified in Section (3); and (ii) when *entendre* embeds a infinitive with its logical subject, the embedded structure must be bigger than VP.

The discussion so far shows that the types of syntactic structures that perception verbs in general combine with do not have to be uniform. In other words, setting aside the particular conceptual restrictions that these verbs may impose on what they combine with, the fact that one type of perception verb attaches at a particular syntactic projection does not mean that another one would attach at the same level. The question that arises is why the two verbs exhibit the observed syntactic asymmetries in the first place. In Section 3, we propose a semantic account of this contrast. In a nutshell, *entendre* can combine with events, and can thus attach as low as VP, which we take to denote a predicate of events, whereas *voir* requires situations, which we take to be introduced above vP, once the event domain has been closed off. These are only the lowest attachment sites. Both verbs can also combine with larger structures as long as their respective semantic requirements are met.[3]

factual observation that sentences like (4b) lack implicit subjects altogether in their embedded clauses. See Section 3.2 for a crosslinguistic comparison.

[3] These requirements concern the direct perception uses of these verbs. For instance, for *entendre* on its direct perception use, the relevant attachment domain ranges from VP up to vP, as only these projections satisfy its semantic requirement. This raises the question of how *que* 'that'-clauses combine with it. We come to this issue in Section 3.1.

3 Semantics of visual and auditory perception

One might attempt to derive the contrast in (2) from a conceptual/ontological difference between the two perceptual modalities. Whereas hearing an event requires access only to the sound that it produces, seeing an event might require visual access to its agent or cause (see for example Raffy & Cyrino 2025). However, we argue that this intuition does not explain the linguistic contrast.

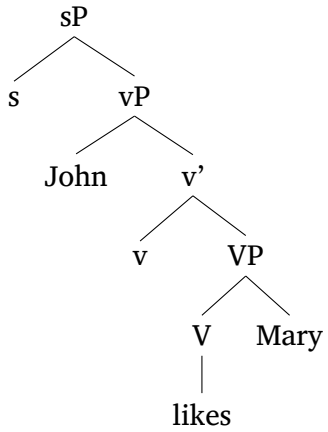
First, its empirical premise is questionable, since one may visually perceive an event without perceiving its initiator or cause. For instance, one may see a door open without seeing who opened it, or see the leaves of a tree move without seeing what caused their movement.

Second, and more importantly, even if visual perception normally required access to an event's agent or cause, it would not follow that this participant must be linguistically expressed in the complement of *voir*. In Neo-Davidsonian event semantics, verbal predicates denote properties of events and may provide only partial descriptions of them. The absence of an agent relation from an event description therefore does not entail that the event itself has no agent. It only means that that the agent is not specified by that description. Consequently, we believe that a condition on what must be present in a visually perceived situation cannot derive a syntactic ban on subjectless event descriptions, as such a ban is actually compatible with a reduced event description.

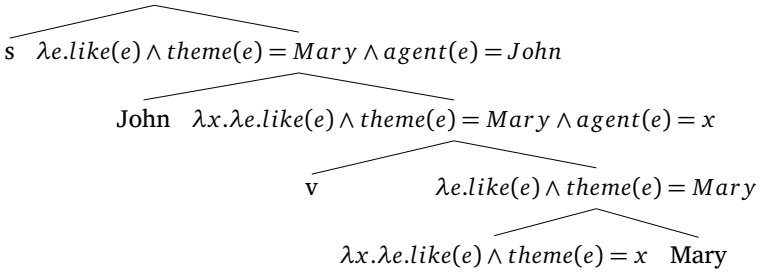
Instead, our proposal is couched within a layered ontology in which syntactic levels correspond to different types of semantic objects; namely events, situations and worlds. We take vP (or however one may call it, e.g., VoiceP Kratzer (1996)) to be the edge of the event domain in the spirit of Ramchand & Svenonius (2014); Ramchand (2018). We further assume that above the event domain, there is a projection, call sP, introducing situations (Ramchand 2018), and closing off the event domain. We do not commit to a particular analysis of what projection must introduce situations. A simple illustration is provided in (6), where sP takes the characteristic function of a set of events and situates it as part of a situation, as implemented in Kratzer (2023).

(6) a. John likes Mary.

b.



c. $\lambda s.\exists e : like(e) \wedge theme(e) = Mary \wedge agent(e) = John \wedge$ e is part of s



d. $[\![v]\!] = \lambda f_{\langle v,t\rangle}.\lambda x.\lambda e.f(e) = 1 \wedge agent(e) = x$

e. $[\![s]\!] = \lambda f_{\langle v,t\rangle}.\lambda s.\exists e : f(e) = 1 \wedge$ e is part of s

We propose that *voir* and *entendre* differ in the kinds of semantic objects that they can directly combine with, and that this semantic difference restricts where they can merge, in the spirit of Wurmbrand & Lohninger (2023), where the syntactic attachment is affected by semantic selection.[4] In particular, *voir* selects situations (type $\langle s, t\rangle$), introduced by sP (e.g., Barwise 1981; Barwise & Perry 1981; 1983). Its combination with the ECM constructions is felicitous as long as the embedded clause denotes situations. By contrast, *entendre* combines directly with predicate of events and consequently attaches within the event domain, as low as VP. Its combination with the embedded ECM material is felicitous as long as the embedded clause denotes events. Not surprisingly, since the vP level denotes a predicate of events,then *entendre* can directly combine with vPs and their subjects.[5]

The idea that *voir* combines with situations is supported by the availability of paraphrases of the type in (7b) for (7a), as noted in previous work (Barwise & Perry 1983; Liefke 2024). Situation-based paraphrases are degraded with *entendre*, as shown in (8b) for (8a), whereas sound-based paraphrases are acceptable. In particular, the objects of *entendre* must correspond to sounds, as understood from the contrast between (8b) and (8c).[6]

(7) a. J'ai vu Nicolas dormir.
I.have seen Nicolas sleep.INF
'I saw Nicolas sleeping.'

b. J'ai vu la situation où Nicolas dormait.
I.have seen the situation where Nicolas slept
'I saw the situation in which Nicolas was sleeping.'

(8) a. J'ai entendu Marie ronfler.
I.have heard Marie snore.INF
'I heard Mary snoring.'

[4] Although we adopt Wurmbrand & Lohninger (2023)'s general idea that semantic selection broadly governs syntactic attachment, we do not commit to their ontology of situations, nor do we commit to their tests to differentiate situation complements from event complements, which do not straightforwardly extend to perception verbs in any case. In particular, while what they characterize as situation complements are events with time and world information, our use of situations is adopted from the situation semantic tradition, where situations are partial entities serving as parts of complete worlds that can themselves act as objects of perception and may contain events as parts (e.g., Barwise & Perry 1983; Kratzer 2023). Thus, in an example such as (7a), our claim is that the subject of *voir* perceives a situation in which an event takes place rather than perceiving a temporally specified event.

[5] Notice that this description does not make any claim as to how these verbs combine with other types of constructions. See Section 3.1 for a discussion.

[6] We thank one of the anonymous reviewers to suggest this rephrase for the previous version of this paragraph.

b. #J'ai entendu la situation où Marie ronflait.
I.have heard the situation where Marie snored
Lit: 'I heard the situation where Mary was snoring.'
c. J'ai entendu le bruit de Marie qui ronflait (≈ j'ai entendu le ronflement de Marie).
I.have heard the noise of Mary who snored/ I.have heard the snore of Marie
'I heard the noise / snoring of Mary.'

The contrasts show us that, unlike *entendre*,*voir* can embed situations. However, as one of the reviewers points out, we have not yet convincingly established that *voir* cannot take events. Establishing this is not straightforward, since, in principle, any nominal that felicitously occurs as the complement of direct perception *voir* can be taken to denote a situation that contains its ordinary denotation (see the discussion in 3.1). Nevertheless, we have identified an event noun that appears to support our claim that *voir* cannot combine with events. Specifically, the noun *pleurs* has two related interpretations. In particular, as a nominalization of the verb *pleurer* 'cry', it may denote the event of crying. Alternatively, it may be synonymous with the noun *larmes* 'tears'. Crucially, while the noun *larmes* is incompatible with event predicates such as *durer* 'last', *pleurs* is perfectly compatible with such predicates, as shown in (9).

(9) a. #Les larmes de Marie ont duré dix minutes.
the tears of Marie have lasted ten minutes.
Int: 'The cryings of Marie have lasted ten minutes'
b. Les pleurs de Marie ont duré dix minutes.
the crying of Marie have lasted ten minutes.
'The cryings of Marie have lasted ten minutes.'

Crucially, *les pleurs de Marie* can occur with *voir* only with the "tear" interpretation. To establish this, we consider a context in which John is behind Mary, who is crying. As shown in (10a), in this context, John can felicitously report that he saw Mary cry, but not that he saw her tears, as the context excludes the latter possibility.

(10) Context: John is behind Marie looking at her, and Marie is crying. John does not see her face.
a. J'ai vu Marie pleurer.
I.have seen Marie cry.INF
'I saw Marie crying.'
b. #J'ai vu les pleurs de Marie.
I.have seen the tears/cries of Marie
Int.: 'I saw Marie crying.'
Available reading (false in this context): 'I saw Marie's tears.'
c. #J'ai vu les larmes de Marie.
I.have seen the tears of Marie
'I saw Marie's tears.'

In particular, since John does not see Marie's face in the context given in (10), any reading in which what John sees is Marie's tears is excluded by the context. This is why (10c) is infelicitous. Now, the question is what John perceives in (10a), which is felicitous in this context. If what is perceived in (10a) were an event, then, given that *pleurs* can denote events, (10b) would be expected to be felicitous on the event reading of *pleurs*.

However, (10b) is infelicitous just like (10c), which indicates that *pleurs* can only receive its 'tears' sense with *voir*. Thus, the contrast between (10a) and (10b) suggests that *voir* does not combine with events, and hence that what John perceives in (10a) is not an event. Importantly, when this nominal is made part of a situation, it becomes felicitous again in the same context, as shown in (11).

(11) J'ai vu une situation dans laquelle il y avait des pleurs.
I.have seen a situation in which there LOC had the cryings
'I saw a situation in which there were cryings.'

As expected, *les pleurs* is compatible *entendre* 'hear', as shown in (12).

(12) J'ai entendu les pleurs de Marie.
I.have heard the tears of Marie
'I heard the tears of Marie.'

Accordingly, we conclude that *voir* takes situations, but not events, as its argument while *entendre* takes events, but not situations. Since projections corresponding to situations become available only once sP is introduced, *voir* must merge at or above sP and is barred from attaching to lower, purely event-denoting positions. *Entendre*, by contrast, combines with events and can therefore attach as soon as an event-denoting predicate is introduced, potentially as low as VP.

Somewhat simplifying, we propose the lexical entries for *entendre* and *voir* as in (13). Consistent with our previous observations in (7) and (8), *voir* takes a predicate of situations and places a situation satisfying the predicate as the theme of the seeing event as in (13a), while *entendre* combines with a predicate of events and places the sound of an event satisfying the predicate as the theme of the hearing event as in (13b).[7]

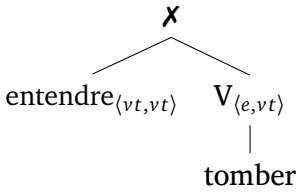
(13) a. $[\![\text{voir}]\!] = \lambda P_{\langle s,t\rangle}.\lambda e.\exists s[see(e) \wedge P(s) \wedge theme(e) = s]$
b. $[\![\text{entendre}]\!] = \lambda f_{\langle v,t\rangle}.\lambda e.\exists e'[hear(e) \wedge f(e') \wedge theme(e) = SOUND(e')]$

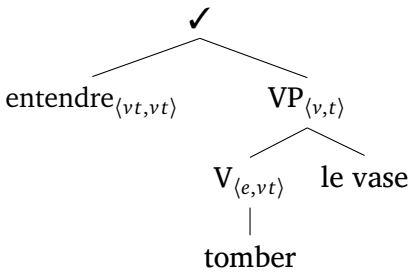
This is already sufficient to account for the contrast observed between *entendre* and *voir* in French. Since *entendre* combines with events, it can attach to a projection as soon as a predicate of events is introduced in that projection. In our current architecture (see (5)), this means that it can attach as soon as the VP is built. For unaccusatives such as (14), this means that *entendre* cannot combine with them unless their internal argument is merged. This is so because unaccusative predicates only denote predicates of events once they combine with an individual as shown in (14c).[8]

[7] The lexical entries in (13) are intended to characterize the direct sensory uses of the two perception verbs in infinitival constructions. See Section 3.1 for a discussion of its implications for other cases.

[8] Our analysis does not make any claim as to the final position of embedded arguments in the structure. As a matter of fact, our representations show the initial merge positions of these arguements, where thematic relations are established and where the type restrictions are satisfied. Our semantic account is compatible with the accounts of ECM where the relevant argument gets case marked without moving to the matrix domain (e.g., Chomsky 1993, see Rosen 1992 for French), or with accounts in which it moves to an object position in the matrix clause, as in (e.g., Ciutescu 2013), albeit possible differences in the size of the embedded clause assumed in these analyses. This is because under the standard treatment, dependencies generated through A movements are interpreted as lambda abstractions over traces and thus do not affect the thematic relations established at the base positions (e.g., Heim & Kratzer 1998). This is why the semantic requirement of these verbs, namely that *voir* combines with situations and *entendre* combines with events, have to be satisfied irrespective of whether the embedded arguments undergo a subsequent A movement. Hence, our account does not need to make a claim as to how the embedded subject receives its case (whether by moving into the matrix domain or by being case marked in situ).

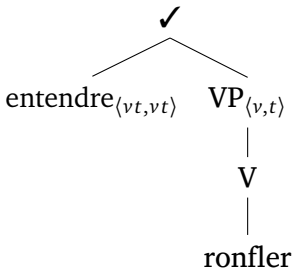
(14) a. Jean a entendu *(le vase) tomber *(le vase).
Jean has heard (the vase) fall.INF (the vase)
'John heard (the vase) falling.'

b.


c.


Unergatives, by contrast, do not have internal arguments, and therefore their bare form, like *ronfler*, already denotes a function of type $\langle v, t\rangle$, which is the right semantic type for *entendre*, hence the grammaticality of (15). Because subjectless unergative constructions embedded under *entendre* involve VP embedding, this also account for the unavailability of agentive modifiers and purpose clauses with such embeddings (see (4b) above), since agents are projected higher at the vP layer.

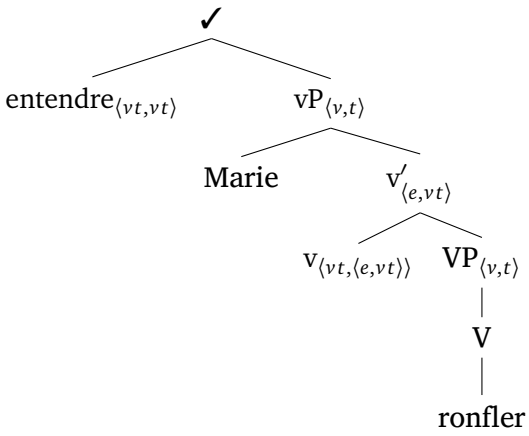
(15) a. Jean a entendu ronfler.
Jean has heard snore.INF
'John heard snoring.'

b.


As we mentioned earlier in Section 2, *entendre* can also embed transitives and unergatives with their overt external argument because they are introduced in [Spec, vP], which has the right semantic type for *entendre*, that is $\langle v, t\rangle$.

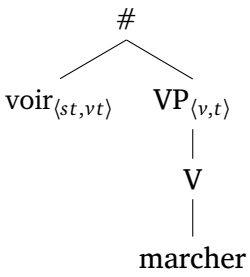
(16) a. Jean a entendu Marie ronfler.
Jean has heard Marie snore.INF
'John heard Marie snoring.'

b.

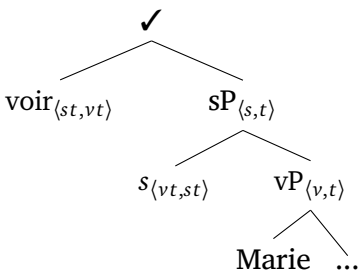


Turning now to *voir*, since it always combines with a predicate of situations, we immediately account for the contrast with respect to the embedding of subjectless unergatives. Recall that *entendre* combines with events, and therefore it can attach at the VP level, which denotes a predicate of events. However, *voir* needs situations, which are introduced after the event domain is closed off. For unergatives, this means that the vP level must already have been projected by the time sP is introduced. Since vP introduces the subjects of unergatives, *voir* cannot embed subjectless unergatives.

(17) a.



b.

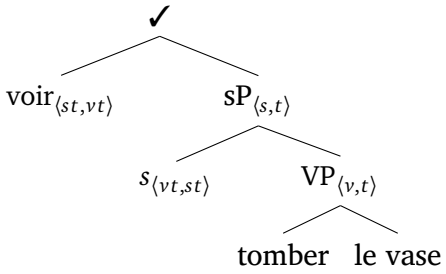


Since these embeddings will necessarily have vPs for unergatives, our account explains why they are compatible with agentive modifiers and purpose (control) clauses (see (4a) above.

Turning now to unaccusatives, those that are felicitously embedded under direct perception *voir*, such as the one in (18), must involve an sP on top of their VP, unlike their counterpart embedded under *entendre* (see (14a)). Since by the time sP is projected, all the required thematic relations are established at the verbal domain, we also account for the unavailability of unaccusative verbs embedded under *voir* without their internal argument, as illustrated before in (3a).[9]

[9] One note is in order. We have adopted the standard assumption that unaccusatives lack a vP layer because they do not introduce external arguments. However, our analysis is compatible with accounts that posit a vP in unaccusative structures, but associate this layer with non-active morphology rather than argument introduction (e.g., Embick 2004). It is also compatible with analyses that decompose the verbal domain into

(18) a. J'ai vu (le vase) tomber (le vase).
I.have seen the vase fall.INF the vase
'I saw the vase falling."

b. 

It is important to emphasize, that in our model, the presence of an agentive level in the embedded material does not mean that the complement is situation denoting. In particular, the embedded constituent could still correspond to vP, and therefore denote a predicate of events. However, the lack of a semantic agent in unergatives entails that the embedded clause denotes an event and not a situation because the presence of an sP would entail the presence of an agentive level for unergatives.[10]

A consequence of our analysis is that French embedded clauses without an overt subject, such as (2a), do not contain an implicit argument either. In this respect, our proposal differs from previous accounts proposed for other Romance languages (e.g., Casalicchio (2016)). This difference, in turn, makes a specific prediction regarding the distribution of depictive predicates in ECM constructions with perception verbs.

Depictives have been used to diagnose the presence of a syntactic argument (e.g., Legendre 1997). In languages like French, where adjectives must agree in phi-features with their subjects, depictives must also phi-agree with this syntactic argument, as shown in (19).

(19) Paul/Marie a crié affolé/e.
Paul/Marie has screamed panicked.M/F.SG
'Paul/Marie screamed panicked.'

The modifiers in (19) function as depictives rather than as event modifying adverbs. Unlike French adverbs, which do not phi-agree, these modifiers agree in gender and number with the argument that they modify. Crucially, the same depictives can be licensed in the absence of an overt syntactic argument, provided than an argument is implicitly represented, as shown in (20). This is supported by the fact that the phi features of the depictive adjective vary according to the situation described, namely, according to whether the swimmer in (20b) and (20c) is grammatically masculine or feminine.

(20) a. Nager nu/nue est rafraîchissant.
to.swim.INF naked.M/F.SG is refreshing
'To swim naked is refreshing, as a man/woman swimmer.'

a root and a verbalizer, both distinct from the head that introduces external arguments, such as Voice (e.g., Harley 2013).

[10] We thank an anonymous reviewer for asking us to highlight this point.

b. Nager nu/nue pour la première fois m'a donné enormément
to.swim.INF naked.M/F.SG for the first time me = has given a.lot
de plaisir.
of joy
'To swim naked (as a man/woman swimmer) for the first time gave me a lot of joy'

c. Nager nu/nue était rafraîchissant.
to.swim.INF naked.M/F.SG was refreshing
To swim naked was refreshing, as a man/woman swimmer'

French also allows for depictives in ECM clauses of perception verbs, as shown in (21), where the depictive phi-agrees with the embedded subject.

(21) a. Un soir j'ai entendu Marie crier affolée.
One night I.have heard Marie scream.INF panicked.F.SG
'One night I heard Marie screaming panicked.'

b. J'ai vu Marie danser nue.
I have seen Marie dance.INF naked.F.SG
'I have seen Marie dancing naked.'

Importantly, given our proposal that cases in which *entendre* combines with an unergative verb contain no agent capable of licensing a depictive modifier, we predict that such modifiers will be ungrammatical in these small clauses. This prediction is borne out in (22), which contrasts sharply with the examples in (21), where the embedded clause contains the external argument of the verb.

(22) *Un soir j'ai entendu crier affolé(e).
One night I.have heard scream.INF panicked.M/F.SG
Int: 'One night I heard screaming panicked (where the screamer is panicked).'

3.1 Other uses of *voir* and *entendre*

The proposal that *voir* combines with situations, but not events while *entendre* combines with events, but not with situations raises the question of what happens when these verbs combine with *that*-clauses and DPs, as in (23).

(23) a. J'ai vu que Marie est partie.
I.have seen that Marie is left
'I saw that Marie left.'

b. J'ai entendu que Marie est partie.
I.have heard that Marie is left
'I heard that Marie left.'

c. J'ai vu un chat.
I.have seen a cat.
'I saw a cat.'

d. J'ai entendu un chat.
I.have heard a cat.
'I heard a cat.'

Given the scope of this squib, it is not possible to do full justice to the complexity of these constructions in this limited space. Therefore, the goal of the following discussion

is to show that the data in (23) are not necessarily incompatible with our proposal and to illustrate various ways to deal with them.

First, the fact that *entendre* combines with events does not necessarily prevent its combination with *that*-clauses, for *that*-clauses arguably specify the content of an attitude rather than serving as its direct argument (Arsenijević 2009; Moulton 2009; Moltmann 2024). For illustrative purposes, and setting aside the details of implementation, cases like (23b) could be analyzed as involving the hearing of the sound associated with an entity, such as a speech event, that supports the truth of the *that*-clause. On this analysis, *that*-clause does not combine directly with the matrix verb; rather, its relation to the verb is mediated by the directly perceived entity, namely the speech event.

The issue is more challenging for DP complements. However, any relational account of perception verbs would have a difficulty accounting for the selectional flexibility of these verbs. One possibility is to assume that these verbs denote functions that have flexible domains.

Another possibility for *see* is that even when it combines with a DP, the DP actually denotes a situation. In such an account, *voir un chat* means perceiving a situation in which a cat appears or in which there is a cat (e.g., Liefke 2024).

Finally, for cases like *entendre un chat*, one possibility is that *entendre* combines with an entity as long as this entity has an associated sound. This entity could correspond to a regular individual like a cat or events. Hence, we could argue that its combination with events is just one instance of its general property that it combines with any entity which comes with an associated sound. This raises the question of why then *entendre* cannot combine with situations if it has a flexible domain. We believe that *entendre* cannot combine with situations because situations do not have associated sounds, while other entities like events or individuals do. This idea is supported by the contrast between (24a) and (24b). While a description of a sound/noise of a situation being annoying is not felicitous as in (24a), event nouns can easily be related to their sounds as in (24b).[11]

(24) a. #Le son/bruit de la situation où il pleurait était agaçant.
the sound/noise of the situation where he was.crying was annoying
'The sound of the situation in which he was crying was annoying.'
b. Le son/bruit de ses pleurs était agaçant.
the sound/noise of his cryings were annoying
'The sound/noise of his cryings was annoying.'

Thus, direct perception *entendre* cannot directly take situations as its argument because situations lack sounds to be heard.

3.2 Crosslinguistic outlook

We have argued that *voir* in French selects situations, whereas *entendre* takes events. The question is whether this generalization extends to other languages. As in French, paraphrases of these ECM constructions with *see* are acceptable with situations in Spanish and Italian, whereas those with *hear* are not, as shown in (25) and (26), respectively. These data are consistent with our generalization about the semantic requirements of these verbs.[12]

[11] Obviously, situations can contain events that have sounds. But this does not make the sound the sound of the situation.

[12] We note that our Italian informants found (26a) quite 'heavy' but accepted it once accepted them as paraphrases of the corresponding ECM constructions. By contrast, they immediately rejected (26b). We thank Mauro Viganò and Andrea Cattaneo for providing the Italian judgments.

(25) Spanish

a. Vi la situación en la que María roncaba.
I.saw the situation in the that Maria was.snoring
'I saw the situation in which Maria was snoring.'

b. #Oí la situación en la que María roncaba.
I.heard the situation in the that Maria was.snoring
'I heard the situation in which María was snoring.'

(26) Italian

a. Ho visto/vidi la situazione in cui Maria russava.
I.have seen/saw the situation in which Maria was.snoring
'I saw the situation in which Maria was snoring.'

b. #Ho sentito/sentivi la situazione in cui Maria russava.
I.have heard/heard the situation in which Maria was.snoring
'I heard the situation in which María was snoring.'

However, in Italian and Spanish, both of which are pro-drop languages[13], the picture seems somewhat different when it comes to embedding infinitives without their overt subject. In particular, Guasti (1993) observes that Italian sentences such as (27) remain grammatical with a silent pro_{arb}.

(27) Ho visto [pro_{arb} rubare dei libri] PRO nascondendoli sotto il cappotto.
I.have seen [pro_{arb} steal.INF some books] PRO hiding.them under the coat
'I saw that some books were stolen by hiding them under the coat'. (Guasti 1993: 126)

Spanish likewise allows infinitival complements without an overt subject under both verbs, as shown in (28)

(28) a. Oí gritar
I.heard shout.INF
≈ 'I heard someone shout.'

b. Vi pelear ayer.
i.saw fight.INF yesterday
'I saw someone fight yesterday'

Since we have argued that French *hear* and *see* combine with different semantic objects, the availability of a bare infinitive with *see* in Spanish and Italian is unexpected under our account. One possibility would be to posit cross-linguistic lexical difference in the type of objects these verbs combine with. However, we believe that the Spanish and Italian data can also be accounted for within our approach.

The Italian example in (27) is analyzed as containing a silent pro_{arb}, unlike what we observed in French. Similarly, we observe that infinitives embedded under *see* and *hear* in Spanish contain implicit external arguments as they license control clauses (see (29) for *hear* and (30) for *see*), and agentive adverbials (see (31) for *hear* and (32) for *see*).[14]

(29) a. Oí gritar pidiendo socorro.
I.heard shout.INF ask.GER help
≈ 'I heard someone shout while asking for help.'

[13] See Rizzi (1982; 1986) for Italian; Fábregas (2025) and references cited there for Spanish
[14] Antonio Fábregas, personal communication, confirmed by another native speaker.

b. Oí gritar para alertar a la policía.
I.heard shout.INF for alert.INF DOM DEF police
≈ 'I heard (someone) shout in order to alert the police.'

(30) a. Vi hacer señas pidiendo socorro.
I.saw make.INF signs ask.GER help
'I saw someone signing (moving the arms, jumping, etc.) while asking for help'

b. Vi atracar a un anciano para quitarle la cartera.
I.saw rob.INF DOM INDEF elderly.man for take.away.INF = DAT DEF wallet
≈ 'I saw someone rob an elderly man in order to steal his wallet.'

(31) a. Oí gritar desesperadamente.
I.heard shout.INF desperately
≈ 'I heard (someone) shout desperately.'

b. Oí golpear el suelo a posta.
I.heard hit.INF DEF floor on purpose
≈ 'I heard someone deliberately hit the floor.'

(32) Vi entregar voluntariamente los documentos.
I.saw deliver willingly the documents.
'I saw someone deliberately deliver the documents.'

Given the presence of implicit arguments in these constructions, unlike in their French counterparts, the Italian and Spanish data do not constitute counterexamples to our generalization that *see* must combine with situations rather than events. Rather, *see* in these languages still embeds an sP containing a vP which projects an implict external argument.[15] In French, by contrast, pro subjects are impossible in these constructions. The use of a subjectless verb therefore corresponds to a lower eventive domain, which is incompatible with the semantic type of *voir*.

4 Conclusion

We have argued that the contrast between (2a) and (2b) follows from the type of objects that *voir* and *entendre* directly combines with, which in turn affects where these verbs can minimally attach to in the syntactic structure in a given semantic domain. Direct perception *voir* requires its argument to denote a predicate of situations and must therefore merge at or above sP. By contrast, direct perception *entendre* requires its argument to be an event predicate and may consequently merge as low as VP. Given that situations are introduced at sP, *voir* cannot embed an unergative predicate without its external argument being projected, for by the time sP is introduced, vP projections are built for unergatives. By contrast, since *entendre* felicitously combines with events, it can attach even when the external argument of an unergative predicate is not yet projected, namely at VP level.

Our proposal has two main consequences. First, it is a clear illustration of how type restrictions of lexical items can interact with where they can syntactically attach in the

[15] As shown above, the corresponding French examples contain no implicit argument, including pro_{arb}. Whether this is correlated to the pro-drop status of Spanish and Italian remains an open question, as do the types of implicit arguments available in French and the conditions under which they are licensed. The point here is narrower in that we suggest that the apparent ability of perception verbs in Spanish and Italian to combine with embedded clauses lacking overt subjects does not undermine the analysis proposed in the squib.

domain of perception verbs. In particular, the asymmetry between *voir* and *entendre* is not purely a matter of syntactic selection resulting from the idiosyncratic features associated with the two verbs, but is a consequence of which syntactic domains introduce which entities and which perception verbs take what type of objects. Our discussion shows that direct perception verbs do not necessarily relate the perceivers to the same type of entity. In this squib we investigated this within the scope of two perception verbs, namely *voir* and *entendre*. We leave to future work the issue of what type of objects other perception verbs take as semantic arguments.

Second, the preliminary observations from Italian and Spanish suggest that it is possible to investigate cross-linguistic variation in perception complements from the same perspective. Given that part of our analysis is based on ontological assumptions regarding the type of entities that can be heard and seen, we expect similar restrictions to surface in other languages as well. We leave a systematic crosslingustic investigation to future research.

Funding information

This research has been funded, either in full or in part, by the French National Research Agency (ANR) under project IMMAGES ANR-22-FRAL-0005-02.

Acknowledgements

We would like to warmly thank Katerina Palasis and Philippe Del Giudice for early discussions on the data. We are also grateful to Clémentine Raffy, Genoveva Puskás, Alexis Mezín and Pierre Ardisson for their native speaker judgments and for discussions. We thank Antonio Fábregas for discussions on Spanish, and Giuliano Bocci, Andrea Cattaneo, Luisa Seguin, and Mauro Viganò for discussions on Italian. We would also like to thank three anonymous reviewers for their feedback and comments, as well as the handling editor, Gabriel Martínez Vera. Finally, we also thank the audience of Going Romance 2025 for their questions. All remaining errors are ours.

Competing interests

The authors have no competing interests to declare.

References

Arsenijević, Boban. 2009. Clausal complementation as relativization. *Lingua* 119(1). 39–50. https://doi.org/10.1016/j.lingua.2008.08.003.

Barwise, Jon. 1981. Scenes and other situations. *The Journal of Philosophy* 78(7). 369–397. http://www.jstor.org/stable/2026481.

Barwise, Jon & Perry, John. 1981. Situations and attitudes. *The Journal of Philosophy* 78(11). 668–691. http://www.jstor.org/stable/2026578.

Barwise, Jon & Perry, John. 1983. *Situations and attitudes*. Cambridge, MA: MIT Press.

Casalicchio, Jan. 2016. *The use of gerunds and infinitives in perceptive constructions: The effects of a threefold parametric variation in some romance varieties* 53–88. John Benjamins Publishing Company. https://doi.org/10.1075/la.234.03cas.

Chomsky, Noam. 1993. *Lectures on government and binding: The pisa lectures*. Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110884166.
Ciutescu, Elena. 2013. Remarks on the infinitival subject of perception verb complements: Evidence for two syntactic configurations. *Revue Roumaine de Linguistique* 58(3). 299–312.
Dikmen, Furkan & Baunaz, Lena & Palasis, Katerina. 2025. French perception verbs are not equal. Paper presented at Going Romance 2025, Ca' Foscari University of Venice, Italy, December 3–5. https://www.unive.it/web/en/15406/home#c164346.
Embick, David. 2004. Unaccusative syntax and verbal alternations. In Alexiadou, Artemis & Anagnostopoulou, Elena & Everaert, Martin (eds.), *The unaccusativity puzzle: Explorations of the syntax-lexicon interface*, Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199257652.003.0006.
Fábregas, Antonio. 2025. Spanish null subjects: descriptive and analytical issues. *Borealis – An International Journal of Hispanic Linguistics* 14(2). 1–158. https://doi.org/10.7557/1.14.2.8246.
Guasti, Maria. 1993. *Causative and perception verbs: A comparative study*. Rosenberg & Sellier.
Harley, Heidi. 2013. External arguments and the mirror principle: On the distinctness of voice and v. *Lingua* 125. 34–57. https://doi.org/10.1016/j.lingua.2012.09.010. SI: Passive: Components and Composition.
Heim, Irene & Kratzer, Angelika. 1998. *Semantics in generative grammar*. Malden, MA and Oxford: Blackwell.
Kratzer, Angelika. 1996. Severing the external argument from its verb. In Rooryck, Johan & Zaring, Laurie (eds.), *Phrase structure and the lexicon*, 109–137. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-015-8617-7_5.
Kratzer, Angelika. 2023. Situations in Natural Language Semantics. In Zalta, Edward N. & Nodelman, Uri (eds.), *The Stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University Fall 2023 edn.
Legendre, Géraldine. 1997. Secondary predication and functional projections in french. *Natural Language & Linguistic Theory* 15(1). 43–87. https://doi.org/10.1023/A:1005728013370.
Liefke, Kristina. 2024. Experiential attitudes are propositional. *Erkenntnis* 89(1). 293–317. https://doi.org/10.1007/s10670-022-00534-w.
Moltmann, Friederike. 2024. *Objects and attitudes*. Oxford: Oxford University Press.
Moulton, Keir. 2009. *Natural selection and the syntax of clausal complementation*. Amherst, MA: University of Massachusetts Amherst dissertation.
Perlmutter, David M. 1978. Impersonal passives and the unaccusative hypothesis. In *Proceedings of the 4th annual meeting of the Berkeley Linguistics Society*. 157–189. Berkeley, CA: Berkeley Linguistics Society. https://doi.org/10.3765/bls.v4i0.2198.
Raffy, Clémentine & Cyrino, Sonia. 2025. Understanding perception: see- and hear-verbs against embedded negation. Paper presented at Going Romance 2025, Ca' Foscari University of Venice, Italy, December 3–5. https://www.unive.it/web/en/15406/home#c164346.
Ramchand, Gillian & Svenonius, Peter. 2014. Deriving the functional hierarchy. *Language Sciences* 46. 152–174. https://doi.org/10.1016/j.langsci.2014.06.013. New Directions in Universal Grammar.
Ramchand, Gillian Catriona. 2018. *Situations and syntactic structures: Rethinking auxiliaries and order in English*. The MIT Press. https://doi.org/10.7551/mitpress/10911.001.0001.

Rizzi, Luigi. 1982. *Issues in italian syntax*. Berlin, Boston: De Gruyter Mouton. https://doi.org/doi:10.1515/9783110883718.

Rizzi, Luigi. 1986. Null objects in italian and the theory of pro. *Linguistic Inquiry* 17(3). 501–557. http://www.jstor.org/stable/4178501.

Rosen, Sara Thomas. 1992. The case of subjects in the Romance causative 17. 79–113. https://doi.org/10.17161/KWPL.1808.638.

Wurmbrand, Susanne & Lohninger, Magdalena. 2023. An implicational universal in complementation: Theoretical insights and empirical progress. In Hartmann, Jutta M. & Wöllstein, Angelika (eds.), *Propositional arguments in cross-linguistic research: Theoretical and empirical issues*, 183–229. Tübingen: Narr Francke Attempto.