

‘Tu bells ulls tens he no y vols veure’: The grammaticalization of the *hi* clitic with Old Catalan and Romance verbs of perception*

Pau Francesch Sabaté

Universitat Autònoma de Barcelona

 ORCID ID 0000-0001-5716-8902

pau.francesch@uab.cat



How to cite: Francesch Sabaté, Pau. 2025. ‘Tu bells ulls tens he no y vols veure’: The grammaticalization of the *hi* clitic with Old Catalan and Romance verbs of perception. *Isogloss. Open Journal of Romance Linguistics* 11(1)/14, 1-47.
DOI: <https://doi.org/10.5565/rev/isogloss.617>

Abstract

Catalan verbs like *veure-hi* (‘see’) and *sentir-hi* (‘hear’) contain a grammaticalized non-referential oblique *hi* clitic (Fabra 1891; Todolí 2002; Russi 2008; GIEC 2016: § 18.6.4.4a). This article explores the role of the inherent *hi* clitic with a group of intransitive perception verbs in Catalan and, secondarily, its equivalent forms in Romance (Occitan, Aragonese, French, Italian and Neapolitan). Firstly, its diachronic origin and grammaticalization process are addressed (Roberts & Roussou 2003; van Gelderen 2011) using corpus data from Old Catalan, Old Occitan and other varieties. It is argued that its origin can be traced back to an instrumental or cause adjunct in centuries 13th in Old Occitan, and 14th in Old Catalan. Secondly, the functional role of the clitic, its expletive status and the syntactic structure of these predicates are discussed. It is shown that two analyses are needed depending on the stage-level or individual-level status of the predicate (Kratzer 1995; Maienborn 2019). In the first case, the non-referential clitic is identified at logical form with an abstract instrument or cause affecting the ability to perceive (Espinal 2009). In the second case, the clitic acts as a committative/instrumental agreement morpheme.

* The quotation in the title belongs to Jaume Roig’s *Espill* (ca. 1460: v. 6966-6967).

Keywords: perception verbs, grammaticalization, inherent clitics, Old Catalan, Old Romance

1. Introduction

Catalan intransitive structures involving the oblique adverbial clitic *hi* —commonly referred as *locative*— grammaticalized with perception verbs such as *veure* for ‘seeing’ and various synonyms; *sentir* and Old Catalan *oir* and *entendre* for ‘hearing’; *sentir* and *notar* for ‘feeling’ (and rarely for ‘smelling’); and *palpar* and *tocar* for ‘touching’,¹ in addition to the cognition verbs *conèixer* and *entendre* (‘knowing’) when they mean ‘perceiving’ or ‘feeling’, have been analysed as the outcome of an intransitivising effect related to this non-referential inherent *hi* clitic (i-clitic or *i-LOC*, the *i* standing for ‘inherent’).² Thus, according to Fabra (1891, 1918, 1956), Nonell i Mas (1906), Par (1926)³, Solà (1990, 1994) and GIEC (2016:§18.6.4.4a), and Russi (2008) for Italian, the *hi* i-clitic turns a transitive verb such as *veure* (1) into a monovalent verb *veure-hi* (2), the nature of this operation remaining nevertheless unknown. Crucially, following the addition of *hi*, these verbs shift their meaning in the sense that they come to mean ‘having the capacity to perceive or understand in a specific way’.

- (1) Catalan
 El Joan *veu* una esquerda.
 the Joan see.3SG a crack
 ‘Joan sees a crack’.

¹ The status of *palpar-hi* is uncertain. Even though it is recurrently included in the list of intransitive perception verbs with *hi* i-clitic by Fabra (1891: § 119, 1912: § 114, 1918: § 43v, 1956: § 76) and also in Todolí (2002: § 6.5.10.4), it does not appear with this use in the normative dictionary DIEC2 and neither in the normative grammar GIEC (2016: § 21.3.3), which states that “[*sentir-hi*] no s’aplica mai als sentits del tacte i del gust. Amb aquests valors i amb referència als sentits del tacte i del gust, usem les perífrasis *no tenir tacte* i *no sentir el gust*” ([*sentir-hi*] does never apply to tact and taste senses. With these values and with reference to the senses of tact and taste, we use the periphrases *not have tact* and *not feel the taste*’ (own tr.)). Besides, no occurrences of the lemma *palpar* + inherent *hi* are found in the CTILC corpus of Contemporary Catalan. As for *tocar-hi*, it has a cognition meaning equivalent to ‘understand’, ‘know’, ‘get it right’ and ‘be mentally sane’. The other verb for tact *notar-s’hi* is mentioned in Rigau (2002: § 15.3.4.1).

² Unlike cases (2) and (3), Catalan regular referential *hi* clitics corefer with PPs and AdvPs that can be proper locatives (i), but also other types such as manner and instrumental adjuncts (ii), commitatives and, also as AdjPs with predicatives and attributes, among other functions (Todolí 2002).

(i) En aquest restaurant_i la Joana *hi*_i dina sovint.
 in this restaurant the Joana LOC have.lunch.3SG often
 ‘In this restaurant Joana has lunch often’.

(ii) Amb bastonets xinesos_i no *hi*_i sé menjar.
 with sticks chinese not INS know.1SG eat
 ‘With chopsticks I don’t know how to eat’.

³ Nonell i Mas (1906: § 4) labels *hi* as an “immanent” clitic in the sense of ‘inherent’ and Par (1926: § 571) as an expletive element.

- (2) Catalan
 El Joan *hi* *veu* (*una esquerda).
 the Joan i-LOC see.3SG a crack
 ‘Joan can see (*a crack)’.

Apart from this effect, a complete analysis of these structures should also account for the optional presence of a non-reflexive clitic with some of these verbs, i.e. *veure-s’hi* (3) (and also more rarely with *oir-s’hi* in Old Catalan). However, the syntactic role of this second non-reflexive clitic will not be explored because it is not attested until mid-17th c. (§ 3) and due to extension reasons.⁴

- (3) Catalan
 El Joan no *s’hi* *veu* per la boira.
 the Joan not SE-i-LOC see.3SG for the fog
 ‘Joan can’t see due to the fog’.

The main aim of this paper is to shed light on the role of the *hi* i-clitic in intransitive structures with the verbs presented—and to determine whether and how it is related to expletivity and valency reduction (Par 1926:§571; GIEC 2016; Real-Puigdollers 2025)—by means of new data about its grammaticalization path, in addition to a detailed analysis of its first-phase syntax and event structure. Hence, the subgoals of this work are twofold. On the one hand, it seeks to establish the diachronic origin of the structure by contrasting the data available from Old Catalan, Old Occitan and other Romance varieties (§ 3.1). Within this goal, an analysis of the grammaticalization path is put forward in § 3.2. On the other hand, a syntactic analysis of these verbs within the distributed morphology (DM) framework and the role of *hi* is presented in § 4 (Halle & Marantz 1993; Hale & Keyser 2002; Mateu 2002; Real-Puigdollers 2013; Acedo-Matellán 2016; Ausensi 2021; Bigolin 2024; among others). Throughout the paper, the focus is placed on the verb *veure* due to its wider crosslinguistic presence with inherent *hi* and its higher frequency in the corpora.

The hypotheses pursued are, first, that the structure might have originated earlier than centuries 16th and 17th in Old Catalan. If right, this would challenge Coromines (1980:233).⁵ As shown in § 3.1, this hypothesis stems from the fact that previous to the analysis presented, earlier instances from centuries 13-14th had been observed in Old Occitan troubadour texts. Second, that, even though its presence in Old Occitan might have influenced Old Catalan, the existence of the structure in both languages among others comes as a result of the properties of the *hi* clitic combined with perception stative predicates. Third, regarding the grammaticalization path, that either an overt prepositional structure *veure en* (‘see in’) + DP was reanalysed as *veure-hi*, or the clitic originally referred to a causal or instrumental adjunct that expresses a condition affecting perception. Fourth, also in connection with its grammaticalization, that the structure *veure* + *hi* firstly appeared with stage-level (SL) predicates, compared

⁴ The non-reflexive clitic is glossed as *SE* regardless of its person, number and proclitic/enclitic position variation: 1st SG *em/-me*, 2nd SG *et/-te*, 3rd SG *es/-se*, 1st PL *ens/-nos*, 2nd PL *us/-vos*, 3rd PL *es/-se*.

⁵ Coromines (1980: 233) observes: “les notícies més antigues que n’he trobat són dels segles XVI i XVII” (“the oldest instances I have found are from centuries 16th and 17th” (own tr.)).

to individual-level (IL) predicates. Fifth, regarding the role of the i-clitic at the syntactic level, it is argued that two *veure-hi* SL and IL structures exist in Old and current Catalan with different dedicated roles of the i-clitic (§ 4). Apart from the hypotheses mentioned, in a recent analysis of *veure-hi*, Real-Puigdollers (2025) has argued that the clitic is the exponent of an expletive pronoun related to a low applicative head (Pylkkänen 2002, 2008), a synchronic proposal that cannot be directly addressed here, but for a remark in § 5.

The methodologies used are, firstly, —to test hypotheses regarding diachrony and grammaticalization— qualitative and quantitative diachronic corpora analyses and, secondly, —to explain the roles of the i-clitic— a syntactic analysis within the DM framework and the Neodavidsonian view on event structure (Davidson 1967; Kim 1969; Maienborn 2019). The main corpora used are the *Corpus informatitzat del català antic* (CICA; Torruella, Pérez Saldanya & Martines 2009) for Old Catalan prose (10–18th c.), the *Corpus informatitzat de la gramàtica del català modern* (CIGCMod; Martines Peres et al. 2024) for Modern Catalan prose (17–19th c.) and the verse corpus *Concordance de l’occitan médiéval* (COM2; Rickerts & Reed 2005) for Old Occitan (17–19th c.).⁶

Following the criteria in the CICA, CIGCMod and COM2, texts are classified in centuries according to the date of the document (original manuscripts, copied manuscripts and prints). The text-type composition of the Catalan corpus data contains, in the CICA, for the 13th c., only prose, mainly from religious literature in *Homilies d’Organyà*, aimed at a parishioner audience, and the hagiography *Vides de Sants Rosselloneses*; secondly, instances from Ramon Lull’s philosophical-theological text *Taula general* (1294), and, third, legal literature from *Costums de Tortosa* (1272) and *Llibre de Cort de Justícia de Cocentaina* (1269–1290). For the 14th c., the sources are more numerous with prose 12 texts belonging to the fields of philosophy (such as Llull’s (1293–) *Lo sisè seny*, religion-theology (such as a 1340 translation of St. Gregori’s *Diàlegs*, law (e.g., *Procés criminal contra Antònia Marquès* (1374–77)), historical chronicle (such as *Llibre dels fets del rei en Jaume* (1343)), medicine (e.g., a translation of Ibn Sina’s *Cànon*) and narrative such as *Questa del sant grasal*. For the 15th c., 16 prose texts belonging to the fields of narrative (mostly, 7 out of 16), religion-theology, medicine, marshalcy, moral and epistolary, and 2 verse texts (March’s poetry and Roig’s *Espill* (ca. 1460). For the 16th c., the corpus basis is scarcer with only 3 prose texts and 1 verse text, all of them about religious matters. Last, in the CIGCMod, for the 17th c., occurrences belong to 6 prose texts form the fields of theatre-comedy (4 out of 7), religion (1), personal diary (1), and 1 verse poetry text. See Table I in the Appendix for a complete account of all occurrences with source, time and genera in the corpora.

From a minimalist viewpoint, the grammaticalization of a clitic into an i-clitic is understood as a process involving the loss or inaccessibility of grammatical *features* and thereby of the grammatical independence of the clitic (Déprez 2003; Roberts & Roussou 2003; van Gelderen 2011; Fischer & Rinke 2013). According to van Gelderen

⁶ The COM2 (Rickerts & Reed 2005) verse corpus was chosen only after searching the prose Old Occitan corpus TMAO (Billy 2017), which contained very few instances of the structures analysed, probably due to its smaller size. (The quantitative data about its word count are not available. They have been requested). In addition to the corpora mentioned, other have been used such as the DICODÒC literary corpus (Dazéas 2025) for Modern Occitan and the corpora in the DMF (Bazin-Tacchella et al. 2015) and BFM (2022) for Old French.

(2011:§2.3, 3.1.2), grammaticalization is driven by Economy Principles such as *feature economy*. In the case of inherent *hi*, unlike referential *hi*, the clitic loses its capacity to corefer with individual entities, events, propositions or situations (generally PPs), because its uninterpretable oblique feature becomes unable to act as probe with PPs. As a consequence, inherent *hi* becomes dependent on intransitive perception verbs like *veure* and *sentir*. In this sense, Espinal (2009) refers to i-clitics in idioms as *lexicalized* elements (see § 4.2).⁷

2. General properties and crosslinguistic presence

Perception verbs have in common that the event they denote involves ability and a presupposed inalienable organ of perception (i.e., eyes, ears, etc.), a feature that distinguishes them from other intransitives, either eventive or stative, with no i-LOC clitic. Beyond this shared property, this section provides a description of other relevant synchronic properties of *veure* and other perception verbs in connection with their event structure (§ 2.1), focusing as well on the SL/IL distinction (§ 2.2), and it also gives an account of their presence with an i-LOC clitic across Romance (§ 2.3).⁸

2.1 The event structure of intransitive perception verbs

As stated by Real-Puigdollers (2025), intransitive perception verbs like *veure-hi*, *sentir-hi* and *palpar-hi* belong to the class of stative eventualities labelled *Kimian states* (KS; Maienborn 2019). KS predicates such as *know*, *weigh* and *see* do not express an event but a state, they are atelic (i.e., they lack an endpoint or a result) and,

⁷ In connection to the grammaticalization of clitics see also Fontana (1993).

⁸ An example of an eventive intransitive also denoting ability is the use of *tallar* ('cut') in (i), while stative intransitives not denoting ability are the Catalan use of *pesar* ('weigh') in (ii), the Catalan verb *brutejar* ('be dirty') in (iii), and, denoting ability, the Catalan verb *regir* ('lead' or 'command') in the senses of 'something, work properly' or 'someone, having mental capacity' as in (iv).

(i) Catalan

Aquest ganivet *talla* bé.
this knife cut.3SG well
'This knife cuts well'.

(ii) Catalan

El Joan *pesa* molt.
the John weigh.3SG a.lot
'John weighs a lot'.

(iii) Catalan

Aquesta camisa ja *bruteja*.
this shirt already be.dirty.3SG
'This shirt is starting to be/look dirty'.

(iv) Catalan, 'regir' in DDLC

El seu enteniment no *regex* [sic.], ni li *vál*.
the his understanding not lead.3SG nor DAT.3SG be.worth.3SG
'His mind does not function, nor is it worth for him'.

most relevantly, they cannot be located in space—they only have a temporal dimension—as shown in (4-6) and, therefore, KS do not allow locative adjuncts.⁹

(4) Mary has known Latin since she was five / *on the stage.

(5) John weighs 70kg today / *on the chair.

(6) Catalan
El Joan *hi* *veu* avui / *a casa.
the John i-LOC see.3SG today at home
'John sees at home'.

Differently, transitive uses of the same perception verbs do involve an event and therefore can be located in space, either as eventive predicates (7), or as Davidsonian states (DS; Maienborn 2019), this is, stative predicates that involve an event and allow telicity such as (8). Opposite to (4-6), in the examples below, the locative adjuncts *at the Verdi Cinema* and *in this room* are perfectly grammatical.

(7) Catalan
La Maria *ha vist* tot Tarantino al Cinema Verdi.
the Mary has.3SG seen all Tarantino at.the cinema Verdi
'Mary has watched all Tarantino's films at the Verdi Cinema'.

(8) Catalan
Aquest retrat de Dante *ha vist* centenars de conferències en aquesta aula.
this portrait of Dante has.3SG seen hundreds of conferences in this room
'This portrait of Dante overlooked hundreds of conferences in this lecture room'.

It is important to note that locations in (4-6) are to be read as physical spaces where states would hold, and not like temporal frames in the sense of 'when (s)he is on the stage / on the chair / at home'. This is relevant because a property of KS perception predicates with inherent *hi* is that the clitic cannot be coindexed neither with low locative adjuncts such as ("at home") in (6), nor with high frame-setting adjuncts. However, as stated by Maienborn (2002), KS predicates do allow modification by high adjuncts setting a temporal-situational frame, such as 'whenever he is on the Moon' in (9), and so do eventualities with perception verbs (§ 2.2).

(9) Catalan
El Joan pesa 11,5 kg, a la Lluna.
the John weigh.3SG 11,5kg at the Moon
'John weighs 11.5kg on the Moon'.

⁹ Regarding telicity, note that despite happening at Krifka's (1989) "atomic times", stative KS verbs can be forced into dynamic eventive readings, where they acquire telicity and, thus, are no longer KS but events:

- (i) Mary *knew* the answer in Paris. ('learnt')
- (ii) China is starting to *own* the Pacific. ('control or occupy, bit by bit')
- (iii) The patient is beginning to *see*. ('gaining sight')

In addition to location in space and telicity, a further property that allows distinguishing a KS is direct observation. Following Maienborn (2019), KS are “reified mental objects” and thus they are not directly observable. In the same way that one cannot see someone know French, internal abilities of seeing, hearing and feeling things cannot be seen. In fact, perception and cognition states can only be observed *inferentially*, i.e., as an implicature, such as in the following Catalan excerpt from the Bible, where, while events like speaking and walking are known from direct evidence, the state of having sight could only be inferred from the blinds’ behaviour.

(10) Catalan, BCI (Mateu 15.31)

Les multitudes s’admiraven veient parlar els muts,
 the multitudes REF-admire.3PL.PST seeing talk the mute
 els mutilats guarits, els coixos caminar i els cecs *veure-hi*.
 the mutilated healed the lame walk and the blind see-i-LOC
 ‘The crowds marvelled at seeing the mute talking, the mutilated healed, the lame walking and the blind seeing’.

2.2 The SL/IL distinction

Parallel to the eventive/stative distinction, event structure can also be characterized through the SL/IL variable (Carlson 1977; Kratzer 1995). According to Maienborn (2019), “SL predicates express (more or less) temporary or accidental properties” such as *being tired, drunk, available, to speak, wait, arrive*, etc., whereas IL predicates “express (more or less) permanent or inherent properties” such as *being intelligent, blond, altruistic, to know French, love, resemble*, etc. Interestingly, intransitive uses of perception verbs like *veure* and *sentir* (*oir* and *entendre* for ‘hearing’ in Old Catalan) allow for both SL and IL readings. As shown in the following pair, even though both sentences equally mean ‘The pilot cannot see’, the exact meaning of the IL predicate (11a) is that the pilot is ‘permanently blind’ (visually impaired), in contrast to the SL predicate (11b), where (s)he ‘cannot see due to a contingent cause’—such as the fog, heavy rain, dizziness or a powerful emotion such as joy or grief—, a temporary state that, either way, is expected to pass.¹⁰

(11) Central Catalan

- a. El pilot no *hi* *veu*. (IL)
 the pilot not i-LOC see.3SG
- b. El pilot no *s’hi* *veu*. (SL)
 the pilot not SE-i-LOC see.3SG

Significantly, in Modern and current Central Catalan, SL cases of *veure* like (11b), have an additional non-reflexive *se* clitic that clusters with *hi* as /si/ (spelled *s’hi*). This morphological difference can be argued to signal the grammatical encoding of the SL/IL pair. However, it must be noted that this morphological distinction is only present in Central Catalan, whereas other varieties such as Valencian have the

¹⁰ Some sources supporting this judgement on the pragmatic distribution of the *veure-hi/veure-s’hi* pair are Nonell i Mas (1906: Ch. XIV, § 1, 193-194), entries for ‘veure-hi’ and ‘veure-s’hi’ in the Catalan normative dictionary DIEC2, and also Real-Puigdollers (2025) who firstly pointed out this contrast with the IL/SL distinction.

grammaticalized *s'hi* cluster in both IL and SL forms of the verb with the i-LOC (*veure-s'hi*) (see ‘veure’ in DNV and § 23.4.2 in GNV).¹¹

As discussed in § 3.2 and § 4, the distinction in meaning between SL and IL predicates plays a crucial role in the grammaticalization process of the i-clitic and also brings evidence for the existence of two different syntactic roles of the clitic. Having said that, Old Catalan SL corpus cases analysed in the following sections do not have the *se* clitic, because its addition is a late development not attested until mid-17th c. in the CIGCMod.¹²

2.3 Crosslinguistic presence

Before exploring the diachronic evolution of the structure in Old Catalan, it is important to note that it is a Romance phenomenon linked to the properties of the locative clitic in association with perception predicates. Besides Catalan, grammaticalized combinations of the verb *see* + i-LOC clitics in intransitive predicates meaning ‘(not) having the ability to see’, have been observed in Aragonese (12), Old and Modern Occitan (13-14), Old and Modern French (15-16) —even though in Contemporary French it is obsolete, dialectal or literary—, Italian (17) and Neapolitan (18).¹³

(12) Aragonese, Ballarín (1971:204)

A escúres no *s'y* *bede* *góta*.
at dark not SE-i-LOC see.3SG drop
‘In the dark it is not possible to see at all’.

¹¹ I thank Manuel Pérez Saldanya (Universitat de València) for confirming the presence of the clitic cluster *s'hi* in both IL and SL *veure* predicates in Valencian Catalan.

¹² The first Modern Catalan occurrences observed in the CIGCMod belong to Pere Esteve’s (1582-1658) writings:

(i) Modern Catalan, Esteve (1582-1658)

A una xica que tenia tan plens de llaganyes los ulls que
to one girl that have.3SG.PST so full of sleep the eyes that
no *s'hi* *veia*, li *passà* la llengua i la *llepà*,
not SE-i-LOC see.3SG.PST DAT.3SG pass.3SG.PST the tongue and ACC.3SG lick.3SG.PST
i *s'hi* *va* *veure*.
and SE-i-LOC go.3SG.PST see
‘To a girl with her eyes so full of sleep crust that she could not see, he passed his tongue over and licked her, and she saw’.

¹³ In Aragonese the i-LOC clitic *i/ie* also appears in *sentir-ie* meaning having the ability either to ‘hear’ or to ‘smell’ (DOA 2023; see also Nagore (2013: § 24). As for Modern Occitan, see also *se ie veire* / *s'i veire* in Mistral (1878) (“s’y connaitre, etre clairvoyant”). The Neapolitan locative and i-LOC clitic is *ce*, cliticized with the verb in (18). In Italian, the i-clitic *ci* is also found with the hearing verb *sentirci* (*Treccani*, also in Russi 2008: 149-153):

(i) Italian, ‘sentire’ in *Treccani*

Non *ci* *sente* da quell’orecchio.
not i-LOC hear.3SG of that-ear
‘(S)he does not hear from that ear’.

- (13) Old Occitan, ‘veser, veire, vire’ in Olivier (2009 [1396])
 per la lbr. de chandelas guastadas a ladita obra,
 for one pound of candles spent in the.said work
 per so quar non y *vezia* point.
 for this that not i-LOC see.3SG.PST point
 ‘due to one pound of candles spent in that work, due to that (s)he did not see at all’.
- (14) Modern Occitan, Charles Perrault (DICODÒC 2014)
 Aquò’s per *i* *veire* milhor, filhòta.
 this-be.3SG for i-LOC see better girl.DIM.VOC
 ‘This is to see better, little girl’.
- (15) Old French, Guillaume de Machaut (1354; DMF 2015)
 Font a Dieu compter a clingnettes, C’est a dire qu’il
 make.3PL before God count at blindfolded it-be.3SG to say that-he
 n’*i* *voit* goute.
 not-i-LOC see.3SG drop
 ‘[old sins and debts] make one count blindfolded before God. This is to say that he can’t see at all’.
- (16) Modern French, Dorgelès (1919)
 On ne va pourtant pas me mettre en sentinelle tout seul?...
 IND not go.3SG even.though not me put in whatch all alone
 Je n’y *vois* presque pas, surtout la nuit.
 I not-i-LOC see.3SG barely at.all especially the night
 ‘They wouldn’t station me as sentry alone, would they?... I can scarcely see, least of all by night’.
- (17) Italian, KIParla Corpus (2019)
 avete tutti modo di leggere o avete bisogno che proietti il testo? //
 have.2PL all way of read or have.2PL need that project.1SG.SBJ the text
 ditemi qualcosa // proietto il testo? //
 say.2PL.IMP something project.1SG the text
 vediamo il testo // *ci vedete* tutti?
 see.1PL the text i-LOC see.2PL all
 ‘Can you all read or do you need me to project the text? Say something to me. Should I project the text? Let’s see the text. Can you all see?’.
- (18) Modern/Contemporary Neapolitan, Andreoli (1887:770)
 ‘*Vederce*’; avere o esercitare la vista.
 see.i-LOC have or exercise the sight.

Despite its crosslinguistic existence, it is worth noting that the presence of the *i*-clitic is only mandatory in Contemporary Catalan (and perhaps Modern Catalan). No precise data have been found about whether the clitic was or is necessary in the other mentioned varieties, apart from a remark about Italian in *Treccani* stating that the form with *i*-clitic *ci* “is used more frequently (...) especially in expressions belonging to the familiar use” and similarly in Russi (2008:151).

3. Diachrony and grammaticalization path

The focus of this section are the corpus data about the existence of *veure* + inherent *hi* structures in Old Catalan and Old Occitan (§ 3.1) and two hypotheses about its grammaticalization path (§ 3.2). The data presented span centuries 13th to 17th. For the corpus analysis, all intransitive finite and non-finite forms of the verb *veure* were searched exhaustively, including every tense, mood, person and dialectal variation and all possible spellings both for the verb and the clitic.¹⁴ For the quantitative analysis, three intransitive forms are compared: 1) *veure* with no clitic, 2) *veure* with a non-reflexive *se* clitic, and 3) *veure* with inherent *hi*. As for Old Occitan, only forms with inherent *hi* of the intransitive *vezzer* ('see') were searched (also through the full paradigm including all finite and non-finite forms). Cases where clitics are referential were carefully excluded on the basis of the following annotation guidelines:

- a) ruling out explicit expressions (PPs, AdvPs, etc.) coreferring with *hi*.
- b) ruling out inferable or presupposed topics coreferring with *hi* (this involves careful analysis of the textual topic and locative salient referents, among other). Special attention is paid to instances where surrounding verbs appear with the *hi* clitic as well, which indicates the presence of plausible antecedents.
- c) ruling out ambiguities with other words such as:
 - cases where the *i* ('and') conjunction is spelled *hi*.¹⁵
 - *veure* followed by adverbs of quantification that grammaticalized from nouns, such as *gens* ('at all') or *res* ('at all', in Old Catalan), which allow for an object interpretation as 'people' or 'not a thing'.¹⁶
- d) ruling out *veure hi* instances with the alternative eventive meaning 'address a mater', 'look into a topic' or the stative 'have knowledge about a topic' explored in § 3.2.1.

3.1 Frequencies of intransitive *veure* forms in the diachronic corpora

The first instances of intransitive *veure* in the CICA corpus date back to the 13th c. They all appear without a clitic such as *veer* in (19) (13 out of 14 cases in this time

¹⁴ The inflection paradigm tables in the DCVB online are especially helpful to this purpose: <<https://dcvb.iec.cat/tables/VEURE.htm>>.

¹⁵ See for instance Joan Roís de Corella's fragment of *Lo cartoxà*, which was excluded, among others:

(i) Old Catalan, Roís de Corella (ca. 1348 i 1377)

yo era cech *hi* ara tinch bona vista, *hi* veyg bé *hi* mire a vosaltres.
 I be.1SG blind and now have.1SG good sight and/i-LOC see.3SG well and look.1SG to you.PL
 'I was blind and now I have good sight, (and?) I see well and I look at you'.

¹⁶ This is the case in the following fragment, where *gens* does not mean 'at all', but a common old spelling of *gents* 'people':

(i) Old Catalan, *Un Llibre de Cort Reial mallorquí del segle XIV* (14th c.)

D[e]manat qí y hera en la braga, e dix
 ask.3SG.PST who LOC be.3SG.PST in the second.rampart and say.3SG.PST
 q[ue] gens *hi* vey, mas q[ue] era[n] estra[n]geras.
 that people LOC see.3SG.PST but that be.3SG.PST foreigners

segment; 93%), but for the one early observation in (20) with a cluster containing the *se* clitic and an additional partitive *en* clitic possibly coreferring with an implicit instrumental adjunct PP *dels ulls* ('from the eyes'). No further cases with partitives were observed.

(19) Old Catalan, Lull (1294)

e home qui és per amar e conèixer Déu,
and man who be.3SG for love and know god
e uylls qui són per veer.
and eyes who be.3PL for see
'And the man that is to love and know God, and eyes that are to see'.

(20) Old Catalan, *Costums de Tortosa* (1272)

On si per aventura lo catiu aurà los ulls afolatz,
where if by chance the prisoner have.3SG.FUT the eyes impaired
o serà orb, (...) dirà lo venedor que bons ulls e bels
or be.3SG.FUT blind say.3SG.FUT the seller that good eyes and nice
ha e que ben se·n veu.
have.3SG and that well SE-PAR see.3SG
'Where if by chance the prisoner's eyes were impaired, or he were blind, (...) the seller will say that he has good and beautiful eyes and that he can see well from them'.

The first instances with inherent locative are observed in the 14th century, where *veure* forms with no clitic remain the most frequent (55 out of 60; 92%), followed by 3 cases with *hi* such as *hi vehia* in (21) and 2 cases with *es*.¹⁷

(21) Old Catalan, *Procés criminal contra Antònia Marquès* (1374-1377)

¹⁷ Interestingly, both 14th c. utterances with *es* have marked polarity items: the fronted quantifier *poch* ('few') in (i), and the fronted manner deictic / positive polarity adverb (PPA) *sí* ('like this / yes') in (ii).

(i) Old Catalan, *Clams i crims a la València medieval* (1279-1321)

mès-lo en una casa de la dita torre, escura,
put.3SG.PST-ACC.3SG in one house of the said tower dark
en la qual aitant *poch se veu* hom ni pot veure
in the which thus few SE see.3SG man nor can.3SG see
de dia com de nit sens lum.
of day as of night without light
'[He] put him in a room of that tower, dark, in which thus anyone hardly sees nor is able to see either in the daytime or night without light'.

(ii) Old Catalan, *Questa del sant grasal* (14th c.)

Cant él fo dadins, lo rey, que tant evia que avie perduda
when he be.3SG.PST inside the king that much have.3SG.PST that have.3SG.PST lost
sa viste e lo poder del cos, *sí·s viu* (...) e ach cobrade
his sight and the power of.the body PPA-SE see.3SG.PST and have.3SG.PST recovered
la viste, e asò fo per volentat de Nostre Senyor.
the sight and this be.3SG.PST for will of our lord
'When he was inside, the king, who had that much that he had lost sight and body strength, *did* see and had recovered sight, and this was by Our Lord's will'.

E ell testis comensà-la de legir e per tal com no
 and he witness began.3SG-ACC.3SG to read and for that as not
hi vehia bé, que f[ò]scant era, no le sabia legir.
 i-LOC see.3SG.PST well that darkening be.3SG.PST not ACC.3SG know.3SG.PST read
 ‘And the witness began reading it [the letter] and since he couldn’t see well
 because it was dusk, he couldn’t read it’.¹⁸

A significant shift is observed in the 15th century, with a spike in cases with inherent *hi*. In this time segment they amount to 17 of 51 cases (33%). The most frequent form remains the absence of a clitic (26 cases; 51%) and the least common is *es* clitic (8 cases; 16%). As for the 16th century, due to the reduction in the Catalan written production the data available for this time segment —4 observations— is too scarce to make any judgements. Having said that, during that period the trend can be assumed to have been positive for the *hi* cases —and perhaps even a tipping point—, since in the 17th century cases with *hi* are already the most frequent with 10 out of 12 observations (83%) and only 2 cases with no clitic, and zero cases with *es*. Thus, between centuries 15th and 17th the “no clitic” form dropped 42% and the form with *hi* became predominant. The fact that the 2 cases with no clitic from the 17th century are marked as belonging to the Valencian variety in the CIGCMod, might indicate that even though the grammaticalization process was complete, the *hi* clitic could have started decreasing in frequency due to the influence of Spanish in the area.

Regarding the *es* clitic form, two remarks are in order. First, the results are unexpected because according to Coromines (1980:233), Old Catalan had “generally” the *se* form and “more rarely some medieval cases of *veure* without *hi* or *se*”.¹⁹ Contrary to this view, the least frequent form since the 13th century is with the *se* clitic, and the most frequent until the 16-17th century is the “no clitic” form. Second, it is interesting to note that, opposite to the other forms, the *se* clitic may allow for a change of state reading in rare cases like (22), which is uncommon among these stative predicates:

- (22) Old Catalan, *Quaresma de sant Vicent Ferrer predicada a València l’any 1413* (15th c.; also cited in Coromines 1980)
 que aquells que no vehen *se* vehen,
 that those that not see.3PL SE.3SG see.3PL
 e los que vehen tornen cechs.
 and the.ones that see.3PL turn blind
 ‘That those who can’t see get sight and those who see become blind’.

¹⁸ A reviewer suggests that in example (21) the *hi* clitic could be a referential locative. However, on the one hand, as seen in § 2.1, intransitive *veure*, as a KS predicate, does not allow for locative adjuncts, and, on the other hand, high frame-setting adjuncts do not pronominalize with *hi*.

¹⁹ “El que es digué generalment a l’E. Mj. és *veure’s*, amb *se* i sense *hi*. (...) més rarament hi ha algun cas medieval de *veure* sense *hi* ni *se*, a la francesa i cast. (...): ‘no veyà dels ulls—’” (‘What was said generally in the Middle Ages is *veure’s*, with *se* and without *hi*. (...) more rarely there are some medieval cases of *veure* without *hi* or *se*, to the French and Castilian style (...): “he did not see from the eyes—” (own tr.)) (Coromines 1980: 233).

See Table 1 for a general count of the absolute and relative (%) frequencies presented up to this point. The last column contains the word total counts of the texts in the CICA (13-16th c.) and CIGCMod (17th c.) for every time segment, which is used to ponder the relative frequencies of each category when comparing different centuries. These can be observed in each category's third column.

Table 1. Absolute and relative frequencies of *veure* + *hi* / *es* / no clitic (13-17th c.)

Century	<i>hi</i>			<i>es</i>			No clitic			TOTAL	
	<i>n.</i>	%	% “ <i>hi</i> ” <i>rel. fr.</i>	<i>n.</i>	%	% “ <i>es</i> ” <i>rel. fr.</i>	<i>n.</i>	%	% “ <i>no cl.</i> ” <i>rel. fr.</i>	<i>n.</i>	<i>word total</i>
13 th	0	0%	0	1	7%	0,00010	13	93%	0,00136	14	959.163
14 th	3	5%	0,00013	2	3%	0,00009	55	92%	0,00243	60	2.258.970
15 th	17	33%	0,00051	8	16%	0,00024	26	51%	0,00077	51	3.360.859
16 th	1	25%	0,00006	1	25%	0,00006	2	50%	0,00012	4	1.623.953
17 th	10	83%	0,00034	0	0%	0	2	17%	0,00007	12	2.914.428
TOTAL	31	22%		12	9%		98	70%		141	

Source: own elaboration.

Regarding textual genera, instances of inherent *hi* are found in the 14th century in prose narratives within historical chronicles and legal literature (the transcript of a trial), where someone's ability to see in a specific context is referred to. In the 15th century, occurrences appear in narrative works, but also in moral-philosophy, medicine-physiology prose —where crucially the ability to see is referred to as an abstract property or sense—, and also a verse poetry text and a verse narrative text. In the 16th century, it is observed in 1 religious verse text. Last, in the 17th century, all instances belong to comedy or religious theatre, but for 1 diary text and 1 religious sermon.

As for the Old Occitan data, the structure with inherent locative *i/hi/y/hy* is found as early as the 13th century in clear examples such as (23). The COM2 corpus contains a total of 17 cases with the verb *vezer*, 10 possibly dating from late 12th c.-13th c. (or early 14th c.), 3 from the 14th c. and 4 from the 15th c. The contexts of the examples were carefully checked for the absence of an explicit or implicit referent for the clitic.

(23) Old Occitan, Uc Brunenc (13th c.)

si no·*i* *vezetz* mentre·l lums es ardens/,
if not-i-LOC see.3PL while-the light is burning
e no·*i* *veiretz*, pois qe·l lums er rescos/.
and not-i-LOC see.3PL.FUT after that-the light be.3SG closed
‘If you don’t see while the light is on, you won’t see after it is turned off’.

For instance, according to the second grammaticalization hypothesis discussed in § 3.2.2, example (24) could be already grammaticalized or on its way, since the instrumental adjunct *dels olhs ab que us remir* serves as a potential referent for the locative clitic *i*. Other ‘see’ synonyms were not searched and cases with *auzir* (‘hear’) and synonyms were not found.

- (24) Old Occitan, Bernat de Bentadorn (13th c.)
 o no *i*? *vei* clar [dels olhs ab que·us remir]_i?
 or not i-LOC see.1SG clear from.the eyes with that-ACC.2SG look.1SG
 ‘Or I don’t see clear from the eyes I admire you with’.

3.2 Two hypotheses on the grammaticalization path

The grammaticalization of the structure is addressed, first, in hypothesis 1 (§ 3.2.1), syntactically by examining a prepositional structure that gave rise to an *hi* clitic with ‘see’ verbs and, secondly, in hypothesis 2 (§ 3.2.2), from an event-structure and pragmatic perspective by analysing the role of the SL/IL distinction, and instrumental and cause adjuncts. The suitability of both hypotheses is contrasted and discussed in § 4 and 5. Last, in § 3.3 some remarks on its grammaticalization in Old Occitan in connection with marked polarity are provided.

3.2.1 Hypothesis 1: the *veer* vs. *veer en* alternation

According to the hypothesis 1, the grammaticalization process of i-LOC clitics in the Old Catalan ‘see’ verb *veser/veer/veure* (among many other morphological and spelling variants) can be traced back to a conative alternation (Levin 1993; Mangialavori Rasia 2021) between the transitive form *veer* (‘see’) and the prepositional intransitive *veer en* (‘see in’). Since the 13th century, an eventive and agentive prepositional verb *veer en* is observed with the closely related meanings ‘to study a matter’, ‘to address a matter’, or more generally ‘to look into’, such as (25). This use is obsolete in Contemporary Catalan and not found in lexicographical works.

- (25) Old Catalan, *Pergamins, processos i cartes reials* 4 (14th c.)
 nós per molts e grans negocis del senyor rey occupats,
 us for many and big business of.the sir king busy
 no podem *veer en* los pleits qui són entre l’Espital de una part
 not can.1PL see in the lawsuit that be.3PL among the-Espital of one part
 e en Berenguer Gomar de l’altra.
 and the Berenguer Gomar of the-other
 ‘Us, being to busy with the king’s business, cannot address the lawsuit...’.

As it can be expected from a PP with the *en* preposition, a referential *hi* clitic corefered with the PP topic that stood for the ‘object’ of work, study or knowledge in these predicates, as seen in (26).²⁰

²⁰ The prepositional form *veer en* is also repeatedly observed in the CICA in the sources *Manual de Consells de la ciutat de València* (14th c.) and *Dotzè llibre del Crestià*, I, by Francesc Eiximenis (14-15th c.). In the first of these sources, the preposition alternates between *en* (‘in’) and *sobre* (‘on’ / ‘about’).

- (26) Old Catalan, *Llibre de Cort del Justicia de Cocentaina* (1269-1290) (13th c.)
 en P(ere) Ferràndeç manà a'n P(ere) Dies, de qui eran
 the Pere Fernàndeç order.3SG.PST to-the Pere Dies of whom be.3PL.PST
 los diners, que no·ls li demanàs
 the money that not-ACC.3PL DAT.3SG request.3SG.PST.SBJ
 entró que él *hi* veés.
 until that he LOC see.3SG.PST.SBJ
 ‘Pere Ferràndeç ordered Pere Dies, who owned the money, not to request it until
 he could address the matter’.

Possibly during centuries 13-14th, this prepositional use underwent a shift from eventive to stative, as a result of a drift in meaning from perception to cognition. As a common property, cognition readings are easily available with perception verbs such as *see* and *hear* due to their inferential and evidential values (i.e., one knows from seeing and hearing). This stative use can be seen in Bernat Metge’s (1399) *Lo somni* (27), among other Old Catalan sources in the CICA. Metge’s example takes place in a chapter addressing the topic of the essence of the human soul, where he ambiguously refers to the “doctors de la Esglesia de Deu” —who are saints that are proclaimed as masters of the faith and doctrine— as people who have *seen in depth*, either in the eventive and agentive sense of ‘delved into’ or the stative sense of ‘known’ about the matter of the soul. How ‘addressing or studying a matter’ becomes ‘knowing about a matter’ comes as a natural consequence: someone who actively addresses the topic of the soul naturally turns into someone who is in the state of being in possession of knowledge about the soul.²¹

- (27) Old Catalan, Metge (1399)
 los doctors de la Esglesia de Deu, los quals molt profundament e be
 the doctors of the church of god the which very deeply and well
hi han vist.
 LOC have.3PL seen
 ‘The doctors of God’s Church, who know’ / ‘have looked into [the matter of the
 soul] very in depth and well’.

This stative meaning is also found in Old Occitan in the COM2 and it exists in current Catalan, as reported by Coromines (1980) in (28), and Italian (29) with an unambiguous stative cognition value.²²

²¹ Apart from Metge’s *Lo somni*, it is observed in the CICA in *La Germania* (14th c.), Jaume Gazull’s, *Lo somni de Johan Johan* (15th c.) and *Epistolari de Ferran I d’Antequera* (15th c.).

²² See DIEC2 (i) and also an example from the first half of the 20th century (ii) (due to the length of the example, instead of glosses, a literal translation is given between quotation marks):

(i) No *hi veig* clar, en aquest afer.
 not LOC see.1SG clear in this matter
 ‘I do not see clear in this matter’.

(ii) Catalan, *Nosaltres sols* (1932)
 Azaña *hi va veure* bé, en els afers de Catalunya: ha estat Francesc Macià, en un míting al Centre Autonomista de Dependents, l’any 1922, qui digué traïdor a Catalunya a Francesc Cambó. Quin sarcasme!... Azaña *hi ha vist* bé. Azaña ha vist que aquesta era l’hora de

- (28) Catalan, Coromines (1980:233; the source of the example is not provided)
 Jo no *hi veig* gens en assò: jo estic del tot a les fosques en assò.
 I not LOC see.1SG at.all in this I be.1SG of.the all in the dark in this
 ‘I don’t see (=know) at all about this: I am completely in the dark about this’.
- (29) Italian, ‘vedere’ in *Treccani*
 in quell’affare non *ci vedo* chiaro, non sono sicuro della bontà
 in that matter not LOC see.1SG clear not be.1SG sure of.the goodness
 o onestà delle intenzioni di chi lo compie.
 or honesty of.the intentions of who ACC.3SG accomplish.3SG
 ‘In that matter I do not see clear (=know), I am not sure of the good will or
 honesty of the intentions of who carries it out’.

During the 14th century, once a stative predicate, at a successive stage the structure underwent a last shift from the meaning ‘having knowledge about a topic’ to ‘having the ability to see’. As explained in § 4, the crucial syntactic feature acquired after its stativization is the transformation into a possession structure, i.e., a possessor individual having knowledge or sight as a possessee. A change in meaning from knowledge to perception could have taken place, given that analogous prepositional structures with the cognition verb *conèixer en* (‘know in’) also underwent changes into perception meanings the other way round, such as the ambiguous a ‘see’ meaning in Old Catalan in (30) and a clear ‘see’ or ‘feel’ meaning in current Catalan forms also with inherent *hi*, and optional non-reflexive *se*, in (31) and (32) (perhaps less frequent but also stated in Espinal’s (2025) dictionary of phraseology).

- (30) Old Catalan, *Cànon d’Avicenna* (14th c.)
 ·l metge que sab conèixer en la fàs del malalt.
 the doctor that know.3SG know in the face of.the ill
 deu guardar en la cara e als vuyls.
 must.3SG look in the face and to.the eyes
 ‘The doctor who knows how to know (=see) into the patient’s face must look
 into the face and the eyes’.
- (31) Catalan, DICPC, Oliveras i Duran (2002)
 Vaig a recuperació i no *m’hi conec* gota.
 go.1SG to recovery and not SE.1SG-i-LOC know.1SG drop
 ‘I go to therapeutic exercise and I feel nothing at all’.
- (32) Catalan, Rigau (2008:§15.3.4.1)
 D’ençà de l’operació *hi he conegut* molt.
 of-since of the-surgery i-LOC have.3SG known a.lot
 ‘Since the surgery I have felt a great difference’.

dominar verdaderament Catalunya. (‘Azaña has seen well, in Catalonia’s affairs: it was Francesc Macià, in a rally at the Centre Autonomista de Dependents, in 1922, who called Francesc Cambó a traitor. What a sarcasm... Azaña has seen well. Azaña has seen that this was the time to truly dominate Catalonia’ (own tr.)).

Having said that, as discussed in § 4.4 and § 5, caution is needed regarding hypothesis 1, because no halfway examples between the cognition and the perception meaning have been observed for *veure-hi* and, therefore, it is not clear how the structure evolved from the meaning in (27) to a genuine perception meaning. Besides this, no analogous conative alternations are found with the Old Catalan verb for hearing *oir* in the CICA corpus. As discussed in § 5, the cognition-like *veure-hi* meaning ‘having knowledge about a topic’ might be a parallel expression forming a subclass with other cognition expressions with inherent *hi*, different from the perception *veure-hi* predicate studied here. Last, as a reviewer points out, it is not clear why the shift in meaning should go from the reading ‘having knowledge about a topic’ to the ‘having the ability to see’, just because both predicates are stative. According to works on grammaticalization such as Traugott (1995) and Hopper & Traugott (2003), the meaning should change from concrete to abstract, from objective to subjective meanings, which is not the case.

3.2.2 Hypothesis 2: the SL/IL distinction and instrumental/cause adjuncts

In § 2.2, a distinction was made between two stative predicates: the SL *veure-hi* denoting a contingent state affecting the capacity to see —such as not being able to see due to the fog or a powerful emotion— and the IL *veure-hi* denoting an inherent state of ability, impairment or effect on someone’s sight. Even though in current Central Catalan these SLs tend to be marked with the *se* clitic, this was not the case in Old Catalan. Thus, these two uses have to be told apart relying on contextual information.

The idea that this distinction is relevant to the grammaticalization process comes from the observation that the first 3 instances of inherent *hi* in the 14th century are SL predicates, whereas IL predicates only appear in the very late 15th century and 8 out of 9 observations belong to the same source *Quesits o perquens*, an anonymous translation from Girolamo Manfredi’s Old Neapolitan *Liber de homine* (1474), printed in Barcelona in 1499. This leads to thinking that the form had been rarely used in IL contexts in the 14th century and probably during the 15th as well. At a different stage, in the 17th century —when the SL form with the *se* + *hi* cluster became grammaticalized— IL cases are more frequently observed (6 instances) than SL cases (4 instances). See Table 2 for a summary of the data (very few observations from the 16th c. were found).²³

²³ Original editions of Manfredi’s (1474) *Liber de homine* were tested to exclude the possibility that the i-clitic was an Old Neapolitan translation calque. However, no equivalent *ce* clitic is found in Manfredi’s source text. Cf. *ued* and *y veu* in the example:

(i) Old Neapolitan, Manfredi (1474)

Perche l’huomo cieco e de piu sotille itellecto che lomo che *ued*.
 why the.man blind be.3SG of more subtle intellect that the.man that see.3SG

(ii) Old Catalan, tr. (1499)

Per què l’ome cech és de més subtil enteniment que l’ome que y *veu*.
 why the.man blind be.3SG of more subtle intellect that the.man that i-LOC see.3SG
 ‘Why has the blind man a more subtle intellect than the man who sees’.

In addition to this, no mention of an i-clitic with perception verbs is found in Ledgeway’s (2009) Old Neapolitan diachronic grammar. Nevertheless, the intransitive form with the *ce* i-clitic *vederçe* appears in Andreoli’s (1887: 770) Moden Neapolitan vocabulary.

Table 2. SL and IL *veure-hi* observations (13-17th c.)

Century	Cases with <i>hi</i>				TOTAL
	IL	%	SL	%	
13 th c.	0	0%	0	0%	0
14 th c.	0	0%	3	100%	3
15 th c.	9	53%	8	47%	17
16 th c.	0	0%	1	100%	1
17 th c.	6	60%	4	40%	10
13 th to 16 th	9	43%	12	57%	21
13 th to 17 th	15	48%	16	52%	31

Source: own elaboration.

Initially, between centuries 14th to 16th, the relevance of the “SL” category as a potential condition for the presence of the i-clitic becomes straightforward after the frequencies of SL predicates are compared for cases with *hi* and “no clitic” cases. As shown in Table 3, SL predicates are relatively more frequent within cases with *hi* (12 out of 21; 57%) than “no clitic” cases (10 out of 83; 12%). The hypothesis of an association between SL uses of *veure* and the presence of inherent *hi* for the values in Table 3 is supported by positive Chi square and Fisher tests with *p* values < 0.00001 and a moderate Cramér’s V of 0.44, indicating a possible association, and also the Odds ratio statistic, that is 10.22 (95% confidence interval: [3.23, 32.35]), which indicates that SL *veure* predicates would be approximately 10.2 times more likely to have *hi* clitics than IL predicates. However, it must be noted that the fact that 8 out of 9 IL observations belong to the same source can be interpreted as a sign of a weaker presence of the i-clitic in these contexts, while at the same time making statistical assumptions less reliable.²⁴

Table 3. SL and IL *veure-hi* vs. *veure* observations (14-16th c.)

Predicate type	<i>hi</i> clitic		No clitic		TOTAL
	<i>n.</i>	%	<i>n.</i>	%	
<i>stage-level</i>	9	43%	73	88%	82
<i>individual-level</i>	12	57%	10	12%	22
TOTAL	21		83		104

Source: own elaboration.

Furthermore, besides proving an association between the presence of the *hi* clitic and SL cases, the nature of this relation should be explained. In § 2.1 it has been shown that, as a KS involving no event, intransitive *veure-hi* cannot be located in space and, therefore, the antecedent of the *hi* clitic could not be a location. Conversely, the view supported here is that SL *veure-hi* predicates are understood as involving a contingent cause or instrument that affects the ability to see, and precisely the expression of that

²⁴ The exact *p* values for the Chi square and Fisher tests are $p \approx 0.0000062$ and $p \approx 0.0001$ ($\approx 1.03 \times 10^{-4}$).

cause might have been the element that the *hi* clitic was coreferential with. Indeed, as it can be seen in (33), *hi* could corefer with a cause or instrumental adjunct such as *ab los ulls ulls mig entelats* ('with the eyes half-blurred'). This example can be assumed to be an ambiguous case on the verge of grammaticalization, where *hi* might still be referential or already inherent.

(33) Old Catalan, Martorell (1490)

a·la cambra de la princessa (...), y Stephania (...);
 at-the chamber of the princes and Stephania
 [ab los ulls mig entelats]_{i?}, scassament *hi*_{i?} podia veure.
 with the eyes half blurred scarcely LOC can.3SG.PST see
 'In the chamber of the princess (...), and Stephania (...); with her eyes half-blurred, hardly could see'.

Besides, it is relevant to note that all intransitive perception predicates involve an organ of perception (eyes, ears, tact in the hands, etc.) such as overt or implicit inherent instrumentals, which might be related to the common grammaticalization path with a clitic that is able to corefer with adjuncts referring to these organs in the Romance languages observed. Apart from providing a link between a particular property of perception verbs and the presence of instrumentals in the event they denote, perception organs involve *inalienable possession* of the instrumental itself, which in turn may be a key factor explaining the presence of the *hi* i-clitic in these predicates contrastingly with other intransitive statives, such as *tallar* ('cut') or *regir* ('function, someone's mind'), also involving ability.²⁵

In addition, a distinction is to be made between external and internal causes. Being more accessible and salient, external causes like a glaring light born of *countless candles*, as exemplified by (33), can be expected to have been the earliest contexts for *veure-hi* uses. At the heart of the matter is the instrumental adjunct *axí* ('like this') italicized in the example. This deictic adverb shows that in cases like (34) the *hi* clitic is towards grammaticalization, understood as dwindling referentiality and feature uninterpretability (van Gelderen 2011; Déprez 2003), albeit antecedents are highly salient in the context and the clitic can still be linked to an explicit textual element.²⁶

(34) Old Catalan, Muntaner (1328)

venien ·II· carretes qui aportaven ·II· ciris, que en cascun avia
 come.3PL.PST 2 carts that come.3PL.PST 2 candles that in each had.3SG.PST
 més de ·X· quintars de cera, qui anaven enceses; (...) altres luminàries
 more of 10 quintals of beeswax that go.3PL.PST lit other illuminations
 eren tantes, que axí bé *hi* podi·om veure con si
 be.3PL.PST many that like.this well i-LOC can.3SG.PST-IND see as if

²⁵ I thank one of the reviewers for stressing the crucial role of inalienable possession of the perception organ to distinguish perception verbs from other intransitive statives without an i-clitic. As noted by the reviewer, if this hypothesis is assumed, then, a relation between instrument and inalienability is to be assumed.

²⁶ I thank Manuel Pérez Saldanya (Universitat de València) for his most valuable suggestions regarding the interpretation of *hi* as initially referring to implicit or explicit cause or instrumental adjuncts.

fos jorn clar.
 be.3SG.SBJ.PST day clear
 ‘There came 2 carts carrying 2 candles, each containing more than 10 quintals of beeswax, that went lit up. (...) lightings were so many that this way one could see as plain as day’.

In other observations also from the 14th century such as (35), the clitic appears in a context where no textual element coreferring with it is explicitly mentioned (e.g., there is no ‘like this’ adjunct referring to the situation), hence, the inherent *hi* can be assumed to be at a more advanced stage of grammaticalization. However, an external cause influencing the ability to see is still remarkably salient, since it is overtly explicated as a parenthetical *que foscant era* (‘that it was dusk’).²⁷

(35) Old Catalan, *Procés criminal contra Antònia Marquès* (1374-1377)

E ell testis comensà-la de legir e per tal com no
 and he witness began.3SG-ACC.3SG to read and for that as not
hi vehia bé, que f[o]scant era, no le sabia legir.
 i-LOC see.3SG.PST well that darkening be.3SG.PST not ACC.3SG know.3SG.PST read
 ‘And the witness began reading it [the letter] and since he couldn’t see well, because it was dusk, he could not read it’.

Secondarily, also during the 14th century, its use might have extended to internal causes like a powerful emotion, such as grief in the early example (36) from *Llibre dels fets del rei en Jaume* (1343). Note that in this case the internal cause is made explicit by the causal adjunct *per lo pesar que n’havem* (‘due to the grief on it’). However, here the clitic is fully lexicalized and is not referentially linked to this causal adjunct, nor to any explicit instrumental adjunct, but to what Espinal (2009:21) labels an “abstract” or “hidden topic” or contextual condition, namely *així* (‘this way’) or *amb aquest pesar* (‘with such a grief’) or other explicatures that might be accommodated in this discourse context (see § 4.2).²⁸

(36) Old Catalan, *Llibre dels fets del rei en Jaume* (1343)

Mas jo só torbat per les paraules e per la hoïda de la sua mort
 but I be.1SG distressed for the words and for the hearing of the his death
 e esta nuyt no y poria veer ne entendre
 and that night not i-LOC can.3SG.PST see nor hear
 per lo pesar que n’havem.
 for the grief that PAR-have.1PL

²⁷ A reviewer points out that it is not clear how causes are related to inalienable possession (of a perception organ or an element that perception depends on), while instruments seem more closer to this notion. See § 5 for a brief comment on the matter.

²⁸ This example was carefully checked with the digitalized manuscript of the oldest edition preserved of *Llibre dels fets del rei en Jaume* (1343). Even though the text is from the 14th century, there are mentions of earlier editions dating between 1276 and 1343, which are lost (Iglesias 2008). Hence, it is not possible to say whether the original text contained the *i*-clitic, which would mean that the phenomenon already existed in the 13th century. In this regard, it is important to notice that this particular chronicle was intendedly conceived with a vivid style close to the spoken language of the time (Ferrando Francés 2024: 10).

‘But I am distressed by the words and the news of his death and that night I couldn’t see, nor hear, due to the grief on it’.

Last, in the 15th century, the use of *veure-hi* expanded to the first IL permanent readings, such as (37), where seeing is a permanent capacity alluded in the context of a contrast between the blind and the ones who see. This step is assumed to take place at a final stage of grammaticalization, where the i-clitic cannot act as probe and, hence, has lost any referential value and independence from the idiomatic expression (van Gelderen 2011; Fischer & Rinke 2013). In these cases, it has been reanalysed to become functionally closer to an agreement morpheme (see § 4.3).

(37) Old Catalan, *Quesits o perquens* (1499)

Lo cech no és ocupat en lo veure, de què la virtut sensitiva e animal
the blind not be.1SG engaged in the see of which the virtue sensitive and animal
no·s diverteix a tants objectes com en l’ome que y veu.
not-SE.3SG deviate at many objects as in the-man that i-LOC see.3SG
‘The blind is not engaged in seeing. This is the reason why the sensitive and animal virtue is not distracted by many objects like in the man who can see’.

From the perspective of the external/internal cause distinction, IL predicates can only fall within the later “internal” category, because they denote an inherent internal property or state. See a summary of this distribution in time segments in Table 4.

Table 4. Predicate type and context

Predicate type	Context	
	<i>external cause</i>	<i>internal cause</i>
<i>stage-level</i>	earlier within 14 th c.	later within 14 th c.
<i>individual-level</i>		last within 15 th c.

Source: own elaboration.

Besides, it is interesting to mention that a successive step that is possible in Contemporary Catalan, but not observed in Old and Modern Catalan, is the presence of the i-clitic in contexts where *veure* is used to refer to ‘sight’ as an abstract property either as an infinitive or a nominalized infinitive (*el veure-hi*; ‘the sight’). These are the only cases in which *hi* is not mandatory in current Central Catalan.²⁹

²⁹ A reviewer suggests that the use of the IL or the SL reading depends on the topic of the texts (e.g., more philosophical or more descriptive content, etc.) and, thereby, proposes comparing IL *veure-hi* cases with alternative expressions used in IL contexts when *veure-hi* does not appear. In this respect, leaving aside the most instances where *veure* simply appears without an i-clitic, the 15th century cases with explicit external causes but with abstract property readings of *veure/veer* never show the *hi* i-clitic. Differently, as shown in (i), the *de* particle or the masculine definite article *lo/el* plus a nominalized infinitive form appears. In addition to this form, the expression *haver mala vista* (‘to have bad sight’) is also found in *Quesits o perquens* (1499):

3.3 Remarks on Old Occitan and marked polarity

As for the grammaticalization of the *i*-clitic with *vezer* in Old Occitan, it is relevant to note that all of the occurrences in the COM2 share a contextual and enunciative property: they can be attributed an expressive value in the sense that they “encode the speaker emotional attitude” (Villalba 2024), in addition to their propositional content. That could be considered rather unsurprising given that gaining or losing sight tend to be dramatic events, perhaps especially in the lyrical register. However, a more objective proof of their expressivity is that 11 out of 17 utterances (65%) contain polarity items, either positive such as the adverbs *be* or *ben* (‘well’), or negative such as the minimizer *guota/gota* (‘not a drop’, meaning ‘not at all’) (38) or coordinate minimizer locutions like *pauc ni pro* (‘not little, nor much’, meaning ‘not at all’) (39) or *ni bas ni aut* (‘not down, nor up’, meaning ‘nowhere’). In contrast, such a strong link with expressivity or marked polarity is not so clearly observed in Old Catalan cases, in which polarity items are found only in 4 out of 31 observations, and only 15 out of 31 cases can be labelled as expressive.

(38) Old Occitan, *Lo Jutgamen General* (15th c.)

elas uelhs avian he *guota* no *hy* *vesian*/
 they.FEM eyes have.3PL.PST and drop not i-LOC see.3PL.PST
 las mas avian he re no toquavan/
 the hands have.3PL.PST and nothing not touch.3PL.PST
 pes avian he no caminavan.
 feet have.3PL.PST and not walk.3PL.PST
 ‘They had eyes and did not see at all; they had hands and they did touch nothing;
 they had feet and did not walk’.

(39) Old Occitan, *Les Mystères provençaux du quinzième siècle* (15th c.)

ha tant gran poder, Quar el retorna als orbz lo *vezer*. (...)
 have.3SG so great power that he return.3SG to.the blind the see
 Hoc, que ieu vigui l’autre dia Hun que, *pauc ni pro*, no *hy* *vesia*,
 yes that I see.1SG.PST the-other day one that little nor much not i-LOC
 see.3SG.PST
 Mas el ly retornet lo *vezer*.
 but he DAT.3SG return.3SG.PST the see
 ‘He has such a great power that he brings back the sight to the blind. Yes, I saw
 the other day one who did not see at all, but he brought him back the sight’.

(i) Old Catalan, Felip de Malla (1483)

yo fos massa sglayat e per açò perdés part *de veer*
 I be.1SG.PST.SBJ too scared and for this lose.1SG.PST.SBJ part of see
 e cert judici, hó per la fuscara de l’ayre.
 and certain judgement or for the darkness of the-air
 ‘I were to scared and due to this I would lose part of the sight and some judgement, or due
 to the darkness of the air’.

4. Syntactic analysis

The grammaticalization paths proposed in § 3 still leave unexplained why the *i*-clitic remained with the verb or, in other words, what its function in the sentence came to be after being reanalysed as a non-referential element, two questions that are addressed here. In this section, an analysis of the event-argument structure is proposed for *veure-hi* and other nontransitive related perception verbs. The first parts of this section (§ 4.1-4.3) explore the syntactic realization of SL and IL *veure-hi* predicates within the instrument adjunct-based hypothesis 2 (§ 3.2.2). In the last section (§ 4.4), the syntactic realization and the benefits of hypothesis 1 are discussed in connection to IL predicates. The following argumentation and the resulting syntactic representation are based on the DM framework and neoconstructionism (Halle & Marantz 1993; Hale & Keyser 2002; Mateu 2002; Real-Puigdollers 2013; Acedo-Matellán 2016; Ausensi 2021; Bigolin 2024).

In a nutshell, DM proposes a minimal set of two relational categories *p* and *v* and a non-relational element *root* ($\sqrt{}$), aimed at decomposing structure into a first-phase syntax and mapping meaning into structural event templates.

- *p* corresponds to a relational *adpositional head* (also named light preposition) here projecting PredP. It introduces a predicative relation among two entities, sometimes akin to a small clause or to a PP but not necessarily (Mateu 2002; Acedo-Matellán 2016).
- *v* corresponds to a *little verb*, projecting *v*P (also named the eventive head, little *v* or light verb).
- $\sqrt{}$ provides “real-world details about the event” (Ausensi 2021:10-11). Roots are non-relational, i.e. they do not project, but may allow merging of other categories such as DP. They are acategorial and need to be categorized via merge operations by “category-assigning functional heads (such as *v*, *n*, and *a*, resulting in the formation of verbs, nouns, and adjectives, respectively)”, but also configurationally (Bigolin 2024).

In addition to these projections, above *v*P, a VoiceP is assumed with the role of introducing the external argument (Kratzer 1996; Acedo-Matellán 2016). Crucially, phonological information and other morphology falling out of the former categories, such as agreement, case marks and, most importantly, *i*-clitics, are assumed to be inserted postsyntactically in a morphology module preceding the phonological form (Embrick & Noyer 2007), alongside with a list of morphological rearrangement operations responsible for the mismatches between the syntactic output and the surface form (Halle & Kesser 2002; Acedo-Matellán 2016). The analysis carried out also builds on the assumptions of the Neodavidsonian turn (Kratzer 1995; Higginbotham & Ramchand 1997; Maienborn 2019), according to which an abstract event projection binding the predication is mapped on syntax, specifically in *v*P.³⁰

I follow Bigolin (2024) in assuming that structure alone captures the division between transitive vs. intransitive eventive predicates and eventive vs. stative KS predicates. Despite involving no Davidsonian event, the syntactic structure of KS still

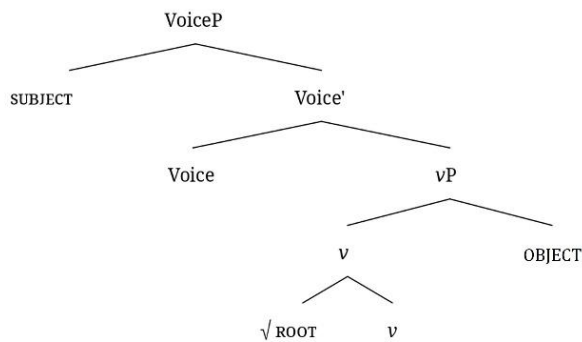
³⁰ No distinction is assumed between types of *v*, depending on the eventive or stative nature of the predicate, but see Acedo-Matellán & Oltra-Massuet (2024).

requires a *v* head capable of verbalizing the acategorical root morpheme. However, as proposed in Bigolin (2024), the position of the subject within *vP* —and specifically within *PredP* in *veure* stative predicates— is what makes a predicate stative (cf. Kratzer 1996; Harley 2017 and Acedo-Matellán & Oltra-Massuet 2024).³¹

4.1 Transitive and intransitive *veure* predicates: perception is a possession

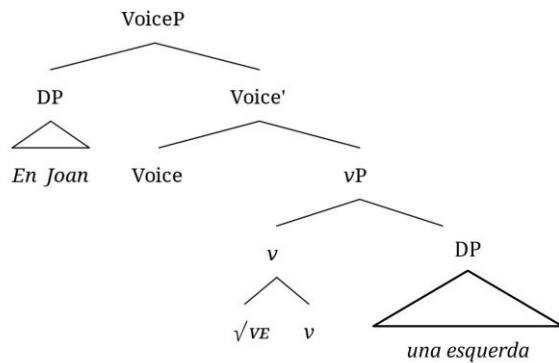
Within DM, a basic transitive structure for a predicate like *veure* involves four functional projections: 1) *VoiceP*; 2) *vP*; 3) the root *VE* (labelled after *veser*, *veer*, *veure*), and 4) a DP object. As illustrated in structure (40) and example (41), *VoiceP* introduces the subject as its specifier; *vP* introduces the event, categorizes the root internally merged with the *v* head and articulates the predication with the object.³²

(40) Eventive transitive structure:



(41) Catalan

En Joan veu una esquerda.³³
 the John see.3SG a crack
 ‘John sees a crack’.



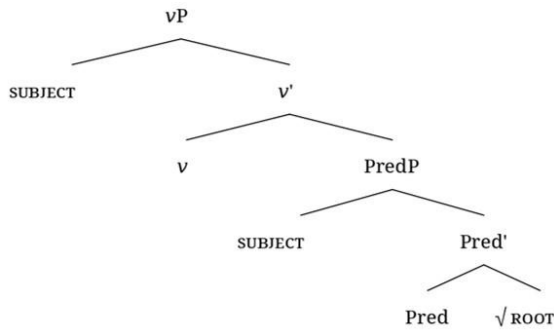
³¹ But see also Ramchand (2008, 2017) for a different approach to stativity.

³² All roots of *veure* are consistently be labelled *VE* despite the varied allomorphy (*vaher*, *vore*, *voure*, *vist*, *vírem*, etc.) and diachronic variation (*veser*, *veer*, *vore*, *voure*, *veure*, etc.).

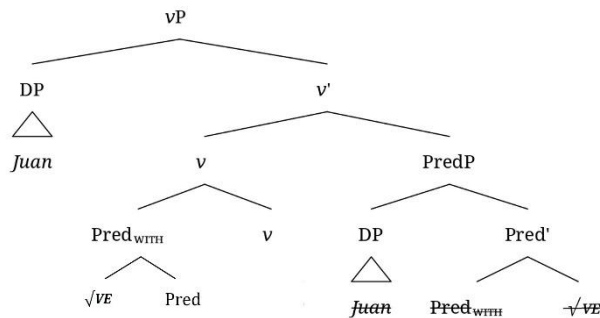
³³ Regarding telicity, within Bigolin’s (2024) proposal, among others, it stems from the kind of object DP, whose features are identified within the predication via the eventive *vP* projection. From this view, an object DP such as *a crack* renders the seeing event an atelic achievement, whereas a DP such as *the football match* introduces a temporal duration dimension making the event telic and consumption-like.

The analysis of the intransitive *veure* has to account for stativity, a feature that must be reflected in structure. Building on Bigolin's (2024) configurational model (which draws on Acedo-Matellán (2016) but differs from his analysis), in this analysis stativity is determined configurationally, because the root does not directly merge with the eventive v head, opposite to the transitive eventive structure in (40-41). Differently, a relational node PredP selects the root as its complement and establishes a relation of predication with the subject in the specifier of PredP. A final low position of the subject at the specifier of the v P will be assumed after movement from [Spec, PredP] to account for the identification of the subject within the eventuality. See the schema of this analysis in (42) and its implementation in a Spanish example—a Romance language without a locative clitic equivalent—in (43), where the *VE* root (of *ver*, 'see') first merges with the Pred head and, subsequently, the Pred head containing the root merges with the eventive v head.

(42) Stative intransitive structure:



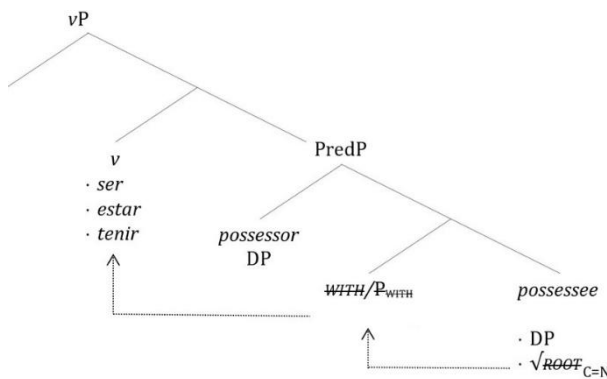
(43) Spanish
 Juan ve.
 Juan see.3SG
 'Juan sees'.



So far nothing has been said on the nature of the Pred head, which is labelled with the *with* subindex in (43). The analysis in this article proposes that intransitive predicates with perception verbs can be equated to the broader group of possession predicates, the key element in common being an explicit or covert accompaniment central coincidence preposition P_{WITH} (Hale & Keyser 2002; Rapoport 2014). According to Kayne (1993), Avelar (2009) and Real-Puigdollers (2013:345), in the broad sense possession predicates are formed by a PredP that relates a possessor subject DP and a possessee DP or root as illustrated in figure (44). The difference

between a DP and a root possessee is that, in [P, Compl], roots share a categorial interpretation with nouns (C=N; Borer 2013) before merging with P_{WITH} and, together with P_{WITH}, with the eventive head *v* to be verbalized, which in turn gives rise to an intransitive structure. In contrast, a DP possessee does not require merge with *v* in sentences with light verbs such as *ser* or *estar* (‘be’) and the transitive *tenir* (‘have’).³⁴

(44) General possession structure:



Central coincidence prepositions relate “two entities in a constant unchanging way”, in most cases in stative predicates (Rapoport 2014:159; also, Hale 1986 and Hale & Keyser 2002). Following Rapoport (2014:160), the *with* preposition is here understood as a functional element in “accompaniment”, “coexistence” or “copresence” relations, but these notions are closely tied to instrument, commitativity, manner and possession in general. Very roughly speaking, the semantics of these stative eventualities of possession can be paraphrased as ‘a possessor (*x*) is *with* possessee (*y*)’, whereby for perception verbs (*y*) is a root. Thus, all the structures encompassed by the schema in (44) share the combination of a light verb and a central coincidence preposition P_{WITH}, which can be explicit, such as in (45), (46) and (47a), or covert, such as in (43), (47b) and (48). The last example with the stative verb *pesar* (‘weigh’), despite being a transitive, is assumed to have a root possession structure (Real-Puigdollers 2013) similar to the stative perception verbs in this research.³⁵

³⁴ I thank Víctor Acedo-Matellán (University of Oxford) for pointing out the relation between perception verbs and possession.

³⁵ According to Avelar (2009), who builds on Kayne (1993), in languages with a verbal expression of possession such as *ter* (‘have’) in Portuguese, the covert preposition *with* merges with the verb and becomes phonologically empty. Languages such as Sango (45), but also Daga (Stassen 2013), Kamaiura (Seki 2000: § 13.2.3) and Egiptian Coptic (Reintges & Lipták 2006), express possession through a light verb and a full *with*-like accompaniment preposition. However, other languages such as Russian (i) and Irish (ii) express possession with full locative *at*-like prepositions:

(i) Russian, Harves & Kayne (2012, cited in Real-Puigdollers 2013)

U menja budet novaja kniga.
at me.GEN will.be new book.NOM
‘I will have a new book’.

(ii) Irish; Harley (2002, cited in Avelar 2009)

Tá an peann ag Maire.
be the pen at Mary
‘Mary has the pen’.

- (45) Sango, Stassen (2013)
 Lo eke *na* bongo.
 3SG be and/with garment
 ‘She has a garment’.
- (46) Middle English, *The Towneley plays* (ca. 1550:Ch.10)
 She was *with* chyld.
 she be.3SG.PST with child
 ‘She was pregnant’.
- (47) Brazilian Portuguese, Avelar (2009:§3)
 a. O Pedro estava *com* documentos da Maria na carteira.
 the Pedro was.3SG with documents of-the Maria in-the wallet
 ‘Peter had Maria’s documents in his wallet’.
 b. O Pedro *tinha* documentos da Maria na carteira.
 the Pedro had.3SG documents of-the Maria in-the wallet
 ‘Pedro had documents of Maria in the wallet’.
- (48) Catalan, Real-Puigdollers (2013:324)
 El Joan *pesa* seixanta quilos.
 the Joan weigh.3SG sixty kilos
 ‘John weighs sixty kilos’.

4.2 The role of instrumental or cause adjuncts in SL *veure-hi* predicates

Having let the general first-phase structure of intransitive perception predicates, it is time to tackle the role of the *i*-clitic. Following the grammaticalization hypothesis 2 (§ 3.2.2), the leading assumption is that the reason for the *hi* clitic before its grammaticalization is a cause or instrumental adjunct that manifests a condition affecting sight, such as *ab los ulls mig entelats* (‘with the eyes half-blurred’ in (33)) or *axí* (‘in this way’, meaning ‘with all that light’ in (34)). In Catalan, the *hi* pronoun can corefer with these adjuncts, despite dialectal and diachronic variation (Todolí 2002: § 6.5.7.2; Rigau 2017, 2020).³⁶

According to Espinal’s (2009:21) work on *i*-clitics, once *hi* is lexicalized as a non-referential element in SL *veure-hi* predicates, the role of the *i*-clitic becomes encoding

λ -abstracted variables which must combine with a (hidden) topic or a (right) dislocated constituent —the argument of a λ -abstract. This topic must not necessarily be explicit in the linguistic context, and might denote some sort of abstract object: a propositional object, a generic situation, a property, an abstract or indeterminate spatio-temporal location, or an indeterminate object.

This is, in this case, an abstracted situational condition that affects the ability to perceive denoted by the predicate. In Espinal’s (2009) view, the identification between

³⁶ In current Valencian and Tortosí varieties the *hi* clitic cannot corefer with cause or comitative adjuncts (Todolí 2002: § 6.5.7.2). Nevertheless, note also that the *hi* clitic is more frequently dropped with perception verbs and in general in these varieties.

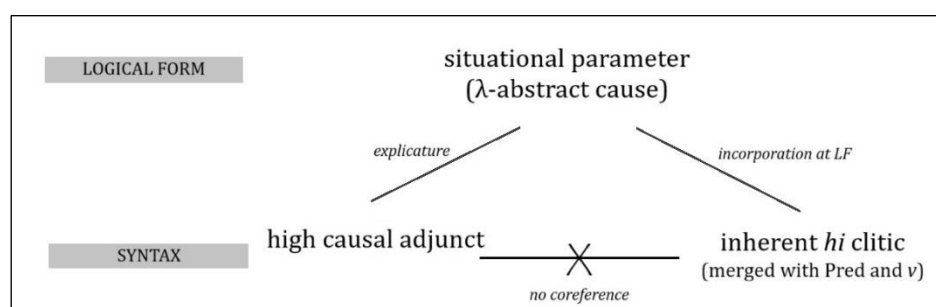
the *hi* clitic and this situational condition does not take place in the syntax, but at logical form (LF) (the conceptual-intentional or semantic interface), given that an *hi* i-clitic is a free variable having only an [+oblique] case selectional feature. As a free variable, it is related to discourse salient information relative to the condition affecting sight. As extensively illustrated in Espinal (2009), non-compositional meanings related to abstract objects are a common phenomenon in idioms with i-clitics in Romance languages and beyond. For instance, in the expression *espifiar-la* ('to mess up'), translatable as 'to put one's foot in it', the accusative feminine singular i-clitic is identified through contextually accessible information with an abstract activity or situation being spoiled or where one messes up, which can be as varied as adding the wrong ingredient to a recipe or blurting out a secret in a conversation. The i-clitic cannot corefer with any explicit antecedent, and if it does, such as in *espifiar-la, una poesia* ('to spoil it, a poem', e.g., by adding a worn-out metaphor), the construction loses its idiomatic value and becomes fully compositional. However, a relevant difference between expressions such as *espifiar-la* and *veure-hi* is that the oblique *hi* i-clitic does not saturate a syntactically required argument of the verb, opposite to the accusative *la*.

Under this formulation, from a syntactic viewpoint, in a SL *veure-hi* predicate with an explicit adjunct like (49), the three elements 1) *hi* i-clitic, 2) overt cause high adjunct *amb la llum de la Lluna* ('with the moonlight') and 3) abstract situational condition relevant in the uttering context "cause affecting sight", would all be related at LF and not in syntax, even though each in different ways, as shown in Figure 1.

(49) Central Catalan

[Amb la llum de la Lluna]_i*, en Joan s'*hi*_i veu.
 with the light of the moon the Joan SE-i-LOC see.3SG
 'By the moonlight, Joan can see'.

Figure 1. Incorporation of the non-referential i-clitic at logical form

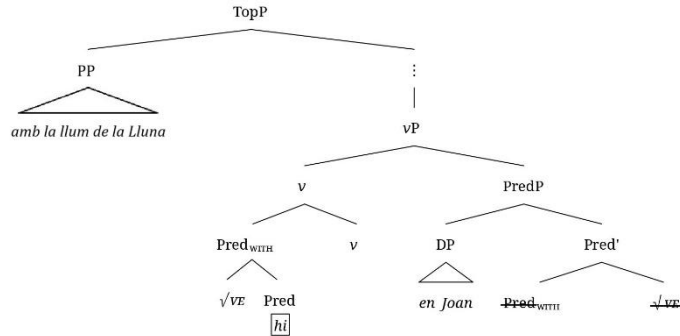


This non-compositional identification at LF with an abstract condition affecting perception in a kind of predicate that involves an inherent organ of perception accounts for the facultative modal layer of their meaning as 'being able to perceive'. In the schema in the figure, while the relationship between the clitic and the abstract condition is one of incorporation at LF, the relation between the high adjunct and the abstract condition is one of explicature, i.e., "a development of logical form" (Sperber

& Wilson 1995:182) or the concrete exemplification of an abstract implicit content which does not necessarily exhaust the range of its possible expressions.³⁷

The syntactic representation of utterance (49) includes the *hi* i-clitic, which is represented as boxed, because it is assumed to be inserted postsyntactically in Pred_{WITH} within the *v* head.

(50) Representation of (49):³⁸



4.3 The function of the i-clitic in IL *veure-hi* predicates

Up to this point, the explanation proposed for the i-clitic relies on the relevance of an abstract cause, a situational condition affecting sight. In the specific case of SL *veure-hi*, this abstract element that Espinal (2009) labels an *empty topic* overlaps with Erteschik-Shir's (2019) notion of *stage-topic*, this is, the implicit or overt "spatio-temporal parameters of the sentence" that form the topic of the predication, which with SL predicates "is the here-and-now of the discourse" (even though Erteschik-Shir mainly provides examples with temporal and locative adjuncts, and not with causes or instruments). However, a relevant insight by Erteschik-Shir (2019:226) is the idea that IL predicates such as *being blond* (51) do not allow the situational stage-topics found with SL predicates like *being in a good mood* (52), because with an IL the subject itself is the stage-topic.³⁹

(51) John is blond #in Cuba / #on Mondays. (IL)

(52) John is in a good mood in Cuba / on Mondays. (SL)

³⁷ In the line of Espinal (2009), to illustrate how the explications of an abstract propositional object or situation linked to a predicate with an i-clitic cannot be exhaustive, one can think of the possible meanings of an idiom with an accusative singular feminine i-clitic like *la in fer-la grossa* ('do it big', meaning 'do something noteworthy'). Imagining that it is uttered in connection to a party, it can be positive, meaning 'throw a great party'; negative, meaning 'a noisy messy party', or an ambiguous judgement, meaning something like a 'great but too expensive party', among the infinity of possible explications.

³⁸ The causal adjunct can be placed in other positions different than [Spec, TopP]. It can be a high adjunct to TP or other layers above vP, depending on the model (see Maienborn 2002).

³⁹ Example (51) can be forced into SL readings whereby 'John dyes his hair whenever he is in Cuba' or 'every Monday', as well as other readings such as 'being just sandy blond in a country where people's hair is darker on average, John stands out as blonder'.

In the same way, IL *veure-hi* readings either are not available with causal adjuncts, or they are infelicitous. In the following examples, the IL readings according to which *my uncle* and *the neighbour* are completely deaf and blind —hence, regardless of the circumstances— are not available. In both cases, the default interpretation is SL: in (53) the uncle has some hearing capacity and in (54) the neighbour has some sight and, thus, their ability to perceive is context dependent.

(53) Central Catalan

Amb el volum tan baix, el meu oncle sord no *hi* sent. (SL/#IL)
 with the volume so low the my uncle deaf not i-LOC hear3.SG
 ‘With the volume so low, the deaf cannot hear’.

(54) Central Catalan

Amb la llum del capvespre, el veí cec no *hi* veu. (SL/#IL)
 with the light of the dusk the neighbour blind not i-LOC see.3SG
 ‘With the light of the dusk, the blind neighbour cannot see’.

This happens because under causal adjuncts, these utterances are forced into SL predicates by default. Otherwise, as ILs, they are not false, but infelicitous utterances containing a causal conventional implicature that cannot hold (Potts 2015), i.e., ‘the affected person does not perceive *because* the stimuli is low’, i.e., the implicated meaning that ‘these people’s perceptive state follows from the cause expressed by the adjunct’ is not true because being completely deaf or blind means having no hearing or sight capacity regardless of the intensity of the volume or light.

Having seen that the role of the i-clitic in SL predicates does not apply in their IL counterpart, it is the moment to explore alternative explanations. The analysis supported here is that, in IL predicates the clitic became reanalysed as an agreement morpheme inserted to *v* that checks a feature of the P_{WITH} head, such as a general [+committative] feature encompassing instrumental-causal and associative-committative values. The clitic would also be inserted postsyntactically in the *v* head with the merged P_{WITH} head. This hypothesis relies on the observation that among the Romance languages with the *veure-hi* i-clitic expression, only Contemporary French does not allow the i-clitic in the *y voir* (‘i-LOC see’) equivalent. This might be due to the fact that the current French locative *y* clitic became more restricted than its equivalents in the other languages observed in § 2.3 and, hence, cannot corefer with most non-locative PPs like instrumental or cause adjuncts such as *ce couteau* (‘this knife’) in (55), opposite to Catalan (Todolí 2002:§6.5.7.2; Rigau 2017, 2020) and Italian (see ‘clitici’ in *Treccani*), and crucially also opposite to Old French, which has committative uses of the *y* clitic, at least when the committative is required by the verb, as in (56), and in Modern French as an adjunct, as shown in (57). Note that Modern French also displayed the i-clitic expression *y voire* (see ‘lourne’ in Greimas 1968:373; also in Grevisse & Goosse 2007:§679a).⁴⁰

⁴⁰ Regarding current Italian, instrumental (i) and committative (ii) uses of the locative *ci* are perfectly grammatical (Paolo Morossi (Universitat Autònoma de Barcelona); PC):

(i) Italian

Io [con la sega]_i ci_i lavoro e ci_i lavorerò sempre piano.

I with the saw LOC work.1SG and LOC work.1SG.FUT always slowly

‘With the saw, I work and will always work slowly’.

- (55) A: On fait quoi [avec ce couteau ?]_i
 INDP do.3SG what with this knife
 B: Je y_i* coupe le pain.
 I LOC cut.1SG the bread
 ‘A: What is this knife for? / B: I cut the bread with it’.
- (56) Cecy sceüz par [eulx-mesmes]_i, car dès le lendemain
 this know.1SG.PAST for they-same since from the tomorrow
 y_i parlay et le veüz à l’œil.
 LOC speak.1SG.PAST and ACC.3SG see.1SG.PAST at the-eye
 ‘I learned this from them directly, for the very next day I spoke with them
 and saw it with my own eyes’.
 (Old French, ca. 1490-1505, Philippe de Commynes, *Mémoires*; BMF corpus)
- (57) de gagner le mal de la reine, en continuant de coucher avec elle_i,
 to acquire the illness of the queen in continuing to sleep with her
 et poussèrent jusqu’à l’inquiéter d’y_i manger.
 and push.3PL.PAST even-to ACC.3SG-worry to-LOC eat
 ‘To contract the Queen’s illness by continuing to sleep with her and they
 even went so far as to bother to eat with her’. (Modern French; Louis de
 Rouvroy, *Mémoires de Saint-Simon*, 1711, Ch. 2; cited in Grevisse & Goosse
 2007:§678)

The fact that Old and Modern French allowed for a more varied use of the locative clitic *y* beyond strictly locative elements (Greimas 1968:338; Moignet 1973:149) might be indicating that a reduction in the selectional features of the oblique *y* clitic associates with its inability to act as an agreement morpheme checking a [+committative] feature of P_{WITH} as an i-clitic in the expression *y voire*. This contrasts with current Catalan and Italian, among other Romance varieties. Such a hypothesis needs confirmation through in-detail research about varieties such as Old and current Occitan dialects, which might accept the i-clitic with intransitive perception verbs while not allowing committative or instrumental uses of the locative clitic or vice versa (see Paradís (forth.)). Research is needed in this field, since the generalization of committative uses of the locative clitic as a frequent phenomenon could be a late development in some Romance varieties, and these committative uses of *hi* might be inconsistent among dialects.⁴¹

Hence, in a nutshell, following the general grammaticalization path “XP > X° > phi-features > zero” in van Gelderen (2011) and Fischer & Rinke (2013), *hi* clitics coreferential with adjunct instrumental (or cause) PPs would have lost its referential DP/PP value (see Rigau 1997:401), first in SL predicates turning into X° i-clitics that

(ii) Italian

Hai visto [Martina]_i? / Sì, ci_i pranzo spesso.
 have.2SG seen Martina / yes LOC have.lunch.1SG often
 ‘Have you seen Martina? / Yes, I have lunch with her often’.

⁴¹ I owe this word of caution to Anna Paradís (University of Oxford; PC), who carries out research on the topic of committatives and clitic climbing in Romance and beyond, and pointed out that the data needed to contrast a hypothesis as such are to be gathered (see Paradís (forth.)).

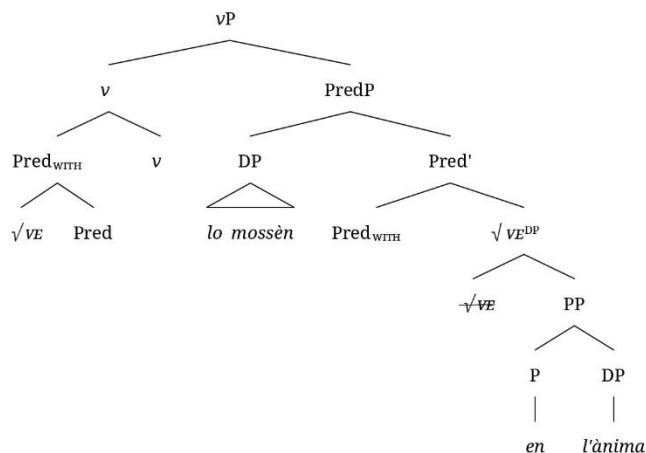
identify at LF with an abstract instrument or cause, and later in IL predicates turning into phi-features as an agreement marker. A final zero-like step of this reanalysis as an agreement morpheme is observed in subvarieties within Valencian and North-Western Catalan, in which the “i-LOC” element lost its clitic nature and became an *i-* morpheme affixed to the infinitive as *iveure’s* (“i-LOC.see-SE”) (Coromines 1980:792; Todolí 2002; Espinal 2009).⁴²

4.4 Benefits and drawbacks of the prepositional structure in hypothesis 1

In § 3.2.1 it has been hypothesized whether the *i*-clitic derives from a prepositional structure *veure en* (‘see in’). Leaving aside for a moment the weaknesses of hypothesis 1 mentioned at the end of § 3.2.1, it is worth to examine how a prepositional antecedent could help explaining the presence of the *hi* clitic. Principally, such an analysis brings in a dedicated feature for the late-insertion of the clitic (Hale & Keyser 2002:78), because the referential *hi* corefers with required PPs headed by *en* —a preposition that assigns [+oblique] case— among other elements (GIEC 2016:§18.6.4; Todolí 2002). However, under this analysis, there are two projections of prepositional nature. In the 14th century stative structure prior to the grammaticalization of the *i*-clitic, aside from the Pred_{WITH} head, the object of *veure* is placed in [P, Compl] and, hence, receives [+oblique] from the *en* (“in”) lower P head, as shown in (58). For the sake of simplicity, example (58) has been made up from the corpus example (25). Note that in this stative possessive structure the root *VE* is categorized as a DP that selects the PP as its complement. This is roughly read as ‘be *with* knowledge *in* the soul’ or ‘be with knowledge in the matter of the soul’.

(58) Old Catalan

Lo mossèn veu en l’ànima.
 the priest see.3SG in-the-soul
 ‘The priest sees into (=knows about) the soul’.

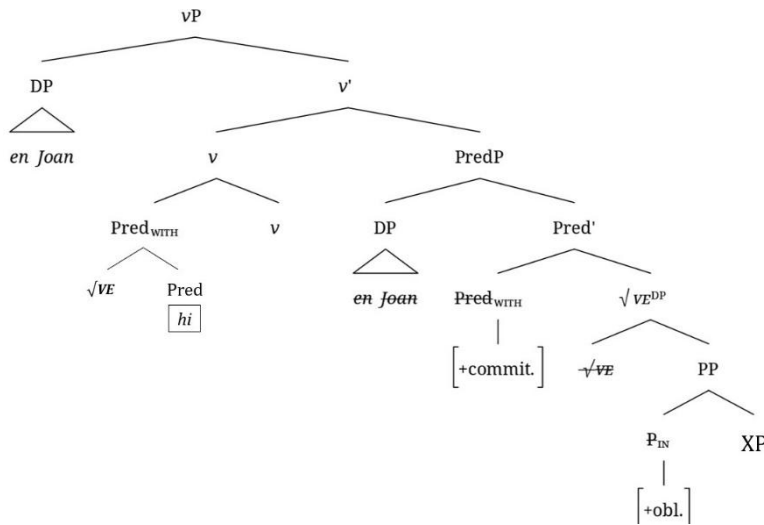


⁴² In these subvarieties, this is observed also in other verbs with inherent *hi*, such as *haver-hi* (‘there is’), becoming *hiaure*, and *valer-s’hi* (‘to be worth/fair/allowed’) becoming *ivaler* (Coromines 1980: 769; Todolí 2002; Espinal 2009).

Consequently, in the IL structure with *i*-clitic, an empty XP is assumed in [P, Compl]. In this way, the non-referential *hi* *i*-clitic is inserted postsyntactically in the v eventive head where $\text{Pred}_{\text{WITH}}$ has already merged. This complex structure would result from a fusion step (Hale & Keyser 2002:78) between two heads $\text{Pred}_{\text{WITH}}$ endowed with a [+committative/instrumental] feature and P_{IN} endowed with an [+oblique] feature, as shown in (59).

(59) Catalan

En Joan *hi* *veu*.
 the Joan *i*-LOC see.3SG
 ‘Joan sees / has sight’.



Despite providing an [+oblique] feature and an antecedent PP for the *hi* clitic, this analysis has varied and significant drawbacks. First of all, as stated in § 3.2.1, no borderline cases between the cognition and the perception *veure-hi* have been observed. Secondly, it adds unwanted complexity and requires positing an empty XP and a complex fusion step. In this respect, there is no sign of how the clitic could have undergone a loss of its ability to act as probe for PPs, thereby shifting from a referential element for the *en* PP into a non-referential *i*-clitic. Third, opposite to hypothesis 2, this analysis does not seem easily extendable to other perception verbs (and perhaps neither to other languages), because there are no observations of an analogous *en* prepositional structure with hearing or tact verbs. Last, regarding a possible relation between the two hypotheses, there is no evidence for an analysis where hypotheses 1 and 2 are understood as historically sequential or compatible, mainly because, as further discussed in § 5, they involve empirically distinct trajectories and there are no corpus observations, nor a theoretical basis, to explain how in this case an object topic PP *en* + DP would reanalyze as a *with*-like PP cause/instrumental adjunct.

5. Discussion

This section is aimed at addressing four research questions that have been left unresolved up to this point: 1) the expletive status of the *hi* *i*-clitic; 2) its assumed

intransitivising function; 3) the plausibility of hypothesis 2 against hypothesis 1, and 4) the applicability of the grammaticalization path of *veure-hi* to the rest of verbs of perception and cognition with an *hi* i-clitic.

Firstly, if expletiveness is to be understood under Tsiakmakis & Espinal's (2022) preliminary definition as "instances of form that have no meaning" or, more generally, with no function, *hi* cannot be considered an expletive because it has proved to carry procedural and functional meanings in SL (§ 3.2.2) and IL predicates (§ 4.3), respectively (a conclusion supported also in Espinal 2009). However, expletiveness can also be better understood as a metalinguistic analytical notion, rather than an actual linguistic category. In this sense, syntactic expletiveness encompasses "functional items (heads or features) that modify another category (either lexical or functional) but do not conceptually contribute to the category they modify" (Tsiakmakis & Espinal 2022). Thus, under this second definition, the i-clitic can be considered an expletive in SL predicates, because even though it modifies the vP, its meaning is only of procedural kind signalling the presence of relevant contextual conditions that affect perception. Differently, in IL predicates *hi* is not an expletive, but an agreement marker related to a [+committative/instrumental] feature.

Second, regarding its intransitivising function (Solà 1990; Todolí 2002; Espinal 2009:§4b; GIEC 2016:§18.6.4.4a), one original conclusion of the grammaticalization analysis supported in § 3.2.2 is that—at least in the case of *veure-hi*—, locative clitics did not (directly) "turn" transitive perception verbs into intransitive forms, but were added to already intransitive stative verbs and grammaticalized into i-clitics. These intransitive predicates did not require a prepositional object either, as explored and argued in § 3.2.1 and § 4.4, but allowed instrumental/cause *with*-like adjuncts, which appeared when *veure* meant 'having sight' or 'the ability to see in a specific way'. The recoverability of these adjuncts by the *hi* clitic serves as an explanatory basis for its grammaticalization into an i-clitic. The fact that their antecedent cause/instrumental adjuncts were not required by the verb seems to go against the idea that the presence of the clitic is related with intransitivity. Nevertheless, the observation that the clitic was added to an already intransitive verb is only a diachronic note, and it does not rule out the possibility that at a later stage the i-clitic became reanalysed as an intransitivising element. In other words, the analysis supported here does not explain why synchronically in Catalan it is not grammatical to use *veure* as an intransitive without the *hi* i-clitic. In addition to this, Real-Puigdoller's (2025) puts forward a synchronic analysis where *hi* plays a valency reduction where it is the exponent of an expletive pronoun complement to a low applicative head. Assuming Real-Puigdoller's analysis, the further research questions are how did a clitic that corefered with a high adjunct and not a required complement grammaticalize, and also how did it reanalyse from a high adjunct into the complement of a low applicative head.⁴³

Third, several arguments have been given against hypothesis 1 (§ 3.2.1) and in favour of hypothesis 2 (§ 3.2.2). The main weakness of the first analysis is that no in-between instances of the shift in meaning from cognition to perception of the prepositional verb *veure hi* have been observed. It is more likely that the PP structure

⁴³ From a different point of view, Espinal (2009: § 4b) proposes that in idiomatic expressions i-clitics perform a semantic argument demotion function. While this looks right for the idioms with accusative and other clitics saturating required thematic roles in the predicates addressed in her work, the case of *veure-hi* requires a different analysis, because diachronically it was used as an intransitive absolute preceding the addition of the clitic.

veure en was an antecedent for a parallel Catalan, Occitan and Italian cognition-related “*veure-hi / i vezer / veder-ci* 2” with a more required “*en*” PP complement antecedent meaning ‘to know about a matter’ before its grammaticalization. In addition, an underlying prepositional structure adds unwanted complexity and an empty XP to the syntax (§ 4.4). Oppositely, the second hypothesis provides an observable referential antecedent—an instrument/cause adjunct—for the diachrony of the *i*-clitic. At the same time, hypothesis 2 establishes a link between the stative KS nature of the predicate and the fact that *hi* could not corefer with a genuine location (§ 2.1), but with a cause/instrument adjunct. Also, it relates the possessive character of the predicate with the P_{WITH} node, which in turn serves as a basis to explain the agreement marking function of the *i*-clitic in IL predicates. Lastly, it hints at a possible explanation for the fact that the *i*-clitic appears specifically with perception verbs and not with other verbs across the Romance languages observed, this is, the presuppositional or intrinsic relation between perception verbs and the unalienable organs of perception (eyes, ears, etc.) as instruments or causes adjuncts in perception predicates. Nevertheless, if unalienable possession of perception organs as instruments in the event is assumed to be the key factor distinguishing predicates with inherent *hi*, then, the syntactic interplay between instruments and inalienable possession is yet to be explained. Another factor that deserves further attention is the distinction between instruments and causes. Since inalienable possession seems to be crucial to the grammaticalization of the *i*-clitic and this variable is hardly compatible with cause—while easily relatable to the notion of instrument—, at some stage of its grammaticalization the presence of inherent *hi* could be expected to have become dependent only on the presence of a relevant instrument. Alternatively, it can be argued that inalienable possession is related to cause as long as a contextual factor affects the implicit perception organs in perception events.

What is more, the instrumental/cause condition helps explaining its ability modal meaning, which would not be encoded in syntax, but derived at LF as an idiom from the identification with the abstract condition affecting perception. As for the relation between the two hypotheses, there is no evidence for an analysis where hypotheses 1 and 2 are understood as historically sequential or even compatible. In this respect, it is important to note that they follow different explanatory models: on the one side, hypothesis 1 is compatible with Traugott’s (1995) and Hopper & Traugott’s (2003) generalisation where grammaticalization goes from concrete to more abstract-functional meanings (i.e., from a concrete-real instrumental adjunct physically affecting perception to an abstract-functional syntactic element), whereas in the other side, hypothesis 2 does not follow this path (i.e., meaning turns from abstract mental cognition ‘know about a topic’ into concrete ‘visual perception in a physical environment’ meaning).

Last, a satisfactory hypothesis should generalize beyond the *veure-hi* case and apply to the other intransitive perception verbs in the analysis. While hypothesis 1 does not seem to explain the *i*-clitic in *oir-hi / sentir-hi* and *palpar-hi* since there are no observations of Old Catalan PP structures *oir en* or *palpar en* in the CICA, hypothesis 2 looks more promising, because cause/instrumental adjuncts can also affect the perceptive capacity in predicates with hearing and feeling verbs. In spite of this observation, it could be argued that other intransitives in the perception-cognition class like *notar-s’hi* (Rigau 2008:§15.3.4.1), *conèixer(-s’)hi* (60) and the particular use of *tocar-hi* in (61) allow identification with a more required abstract object that can be

explicated with *en* PPs, hence, different of cause/instrumental adjuncts. In (60), the abstract object would be “in the effects of the therapy”, and in (61), *tocar-hi* has a left-dislocated object with the *en* preposition that explicates the abstract object “in the topic of stock market”. If this is right, these predicates would form a different subclass alongside with the cognition meaning of *veure-hi* 2 and a syntax like the one hypothesis 1 would better fit them.⁴⁴

(60) Catalan, DICPC, Oliveras i Duran (2002)

Vaig a recuperació i no m'hi conec gota.
go.1SG to recovery and not SE.1SG-i-LOC know.1SG drop
'I go to therapeutic exercise and I feel nothing at all'.

(61) Catalan, 'tocar' in DEIEC

En questions de borsa hi toca.
in matters of stock.market i-LOC touch.3SG
'When it comes to stock market (s)he knows a lot'.

6. Conclusions

This research has explored the role of an *hi* i-clitic with a group of intransitive perception verbs mainly in Catalan and its equivalent forms in other Romance languages (Occitan, Aragonese, Italian, Neapolitan and French). The paper aimed at, firstly, raising new data through a corpus diachronic analysis of the grammaticalization path of this structure focusing on the Old Catalan and Old Occitan verb for seeing *vezer/veer/veure* (among other variants) and, secondly, applying the results to a grammatical analysis of the i-clitic. Following this line of inquiry, in § 3.1 it has been shown that the structure with i-clitic first appears in the 13th century in Old Occitan, and in the 14th century in Old Catalan, a finding that challenges Coromines (1980). The absolute intransitive form without a clitic is the most frequent one between centuries 13th and 16th, whereas the form with an *es* clitic is the least frequent at all times. The tipping point for the *hi* i-clitic form comes in the 17th century, when it becomes widely predominant, even though due to the lack of data from the 16th century it cannot be discarded that the generalisation of the *hi* form took place earlier. As for the clitic cluster *es + hi*, it is not attested until the 17th century.

Crosslinguistically, inalienable possession of an organ of perception has been assumed to be the common feature among perception verbs with i-clitics, which distinguishes them from other intransitives involving ability with no i-clitic (e.g., *tallar* ('cut')). However, in Old Occitan the i-clitic appears mostly in expressive and marked polarity environments (§ 3.3), whereas in Old Catalan the contextual condition triggering it is the presence of an instrument or a cause affecting perception —or explicitly the inherent organs of perception— in SL predicates (Kratzer 1995; Maienborn 2019). As shown in § 3.2.2, initially, this cause tends to manifest as an

⁴⁴ Within this same group, see also French *s'y connaître en* in Grevisse & Goose (2007: § 680):

(i) Ce huron s'y connaît en peinture.
this huron SE-i-LOC know.3SG in painting
'This wretch knows about painting'.

explicit adjunct and has an external nature, such as light or darkness. At a more advanced stage, it can be implicit and have an internal nature, such as a powerful emotion. At a later grammaticalization stage, the presence of the *i*-clitic extends to IL predicates for which perception is an intrinsic capacity, regardless of contextual conditions. The grammaticalization of the inherent clitic has been explained as a loss of the ability to act as probe of its grammatical features (van Gelderen 2011; Déprez 2003). Specifically, its uninterpretable oblique case feature. As a consequence, the clitic lost its referential value and cannot corefer with the mentioned instrumental and causal adjuncts. In turn, this made the *i*-clitic syntactically dependent on intransitive perception verbs. In § 4.4 and 5, it has been argued in favour of this grammaticalization path (labelled hypothesis 2) and against hypothesis 1, according to which the structure derives from an antecedent prepositional structure *veer en* (“perception verb + PP”).

Regarding the syntax of the structure and role of the *i*-clitic, building on Real-Puigdollers (2013) and Maienborn (2019), in § 4.1 it has been proposed that intransitive perception verbs are to be analysed as stative possession structures with a first-phase syntax containing a relational head P_{WITH} that carries an [+instrumental] or a general [+comitative] feature. As seen in § 4.2, following Espinal (2009), in SL predicates the *i*-clitic identifies at LF with an abstracted situational condition that affects the ability to perceive. Also at LF, the identification with this condition accounts for the ability modal meaning ‘(not) being able to see/hear/feel’ usually salient with these structures. Hence, its modal value is not necessarily encoded in the syntactic level. As a free variable, the *i*-clitic is related to discourse salient information relative to the condition affecting perception. Differently, in IL predicates, the *i*-clitic is grammaticalized into an agreement marker that checks the instrumental or comitative feature of the P_{WITH} head.

In light of the data, this analysis challenges the view according to which the *hi* *i*-clitic is an expletive. It cannot be related to expletivity understood in the broad sense (Tsiakmakis & Espinal 2022), since despite not carrying a propositional meaning itself, it has been argued to have a procedural meaning in SL predicates (Espinal 2009) and a role in syntax in IL predicates. Apart from that, the study feeds new data to the open question of the clitic’s intransitivising role. From a diachronic perspective, the *i*-clitic did not directly reduce the valency of perception verb, firstly, because it appears in a context where intransitive *veure* is widely used as an absolute and, secondly, because its function is originally related to the presence of an implicit or overt instrumental/cause adjunct and not with a required argument. This finding leaves the door open for future analyses, such as Real-Puigdollers (2025), of the later reanalysis of the *i*-clitic into an intransitiviser, the element responsible for the ungrammaticality of the intransitive perception verbs without the *hi* *i*-clitic in current Catalan. More research is needed to develop a precise syntactic implementation of feature loss within structures with and without inherent *hi*, and also to explain why *hi* checks for an instrumental feature, possibly in connection with the presupposed nature of perception organs as unalienably possessed instruments with perception verbs. In addition, the syntactic role of the non-reflexive *se* clitic within the structure and the reasons that led to its grammaticalization in Catalan in the 17th c. are yet to be explored, possibly in connection to the analysis of *hi* and *se* in Real-Puigdollers (2025). Another issue that deserves more clarification is the presence of *se* and the degree of optionality/obligatoriness of *hi* and *i*-LOC equivalents across Romance varieties.

As a final remark, more research in the line of Paradís (forth.) is needed in the field of the instrumental, causal, committative and other non-locative functions of the adverbial clitics across Old and contemporary Romance varieties. To this respect, with this kind of data, multiple linguistic assumptions and proposals, such as the analysis in § 4.3 according to which in IL predicates the *i*-clitic is related to a instrumental/committative feature, could be further tested by contrasting the range of the functions of the locative clitic across Catalan, Occitan, French and Italian and Neapolitan among other varieties and its manifestations in grammaticalized structures. Lastly, a fine-grained syntactic account of the relation between the instrument feature and inalienability of perception organs is needed, as well as facultative modality.

Acknowledgements

This research benefited from the project *Expletiveness and expressive meaning* (PID 2023-150347NB-I00) (AEI). I thank Cristina Real-Puigdollers, Manuel Pérez Saldanya, Xavier Villalba, Teresa Espinal, Víctor Acedo-Matellán, Jaume Mateu and Anna Paradís for their helpful comments and guidance.

References

- Acedo-Matellán, Víctor. 2016. *The morphosyntax of transitions*. Oxford: OUP.
- Acedo-Matellán, Víctor & Isabel Oltra-Massuet. 2024. Prepositional Verbs and the Individual-/Stage-Level distinction. *Glossa: a journal of general linguistics* 9(1): 1-48.
- Andreoli, Raffaele. 1887. *Vocabolario napoletano italiano*. Torino, Roma, Milano, Firenze: Paravia.
- Ausensi, Josep. 2021. *The contribution of roots: The division of labor between grammar and the lexicon in meaning composition*. Barcelona: UPF dissertation.
- Avelar, Juanito. 2009. The comitative-copular basis of possessive-existential constructions in Brazilian Portuguese. In: J. Nunes (org.). *Minimalist essays on Brazilian Portuguese syntax*. Amsterdam: John Benjamins, 139–160.
- BFM = *Base de Français Médiéval*. 2022. Lyon: ENS de Lyon, Laboratoire IHRIM. txm.bfm-corpus.org.
- Bigolin, Alessandro. 2024. The syntax of argument structure. Elements for a configurational theory. Ph.D. thesis, Bellaterra: Universitat Autònoma de Barcelona.
- Borer, Hagit. 2013. *Structuring Sense. Volume III: Taking form*. Oxford: Oxford University Press.
- Carlson, Gregory Norman. 1977. Reference to kinds in English. Ph.D. thesis, Amherst: University of Massachusetts.

BCI = Bíblia catalana interconfessional. *Nou testament*, Mateu, 15:31.

CICA = Torruella, Joan (dir), Pérez Saldanya, Manel, & Josep Martines. 2009. *Corpus Informatitzat del Català Antic*. <http://CICA.cat>.

CIGCMod = Martines Peres, Josep; Martines Peres, Vicent; Antolí, Jordi, & Sánchez López, Elena. 2024. *Corpus Informatitzat de la Gramàtica del Català Modern*. <https://metaconcor.kiobus.net/search>.

COM2 = Rickerts, Peter T. (dir.), & Alan Reed. 2005. *Concordance de l'occitan medieval*. [local software].

Coromines, Joan. 1980-2001. *Diccionari etimològic i complementari de la llengua catalana* (= DECat). Barcelona: Curial, La Caixa.

CTILC = Rafel i Fontanals, Joaquim, & Sandra Montserrat (dir.). *Corpus Textual Informatitzat de la Llengua Catalana*. Barcelona: Institut d'Estudis Catalans. <https://ctilc.iec.cat/scripts/index.asp>.

DDLC = Secció Filològica de l'Institut d'Estudis Catalans, & Joaquim Rafel i Fontanals (dir.). 2016. *Diccionari descriptiu de la llengua catalana*. Barcelona: Institut d'Estudis Catalans. <https://dcc.iec.cat/ddlcI/scripts/article.asp>.

Davidson, Donald. 1967. *The logical form of action sentences*. In N. Rescher (ed.), *The logic of decision and action*, 81-95. Pittsburgh: University of Pittsburgh Press.

DCVB = Alcover, Antoni Maria, & Francesc de Borja Moll. (2001-2002 [1975-77]) *Diccionari Català-Valencià-Balear*. Palma. <http://dcvb.iecat.net>.

DEIEC = Secció Filològica de l'Institut d'Estudis Catalans. 2023. *Diccionari Essencial de la Llengua Catalana*. Barcelona: Institut d'Estudis Catalans.

Déprez, Viviane. 2003. Moving phrases in nominal phases. Paper presented at the *LSRL XXXIII*, Bloomington: Indiana University.

DICODÒC = Dazéas, Benoît (dir.). *Diccionari occitan de referència*. (Còrpus literari occitan). Lo Congrès permanent de la lenga occitana. dicodoc.eu/oc/corpus.

DIEC2 = Secció filològica de l'institut d'Estudis Catalans. 2007. *Diccionari de la llengua catalana*. Barcelona, Palma, València: Edicions 3 i 4, Edicions 62, Enciclopèdia Catalana, Moll, Publicacions de l'Abadia de Montserrat.

DMF = Bazin-Tacchella, Sylvie; Martin, Robert, & Gilles Souvay. 2015. *Dictionnaire du Moyen Français (1330-1500)*. zeus.atilf.fr/dmf/smart/.

DOA = Estudio de Filología Aragonesa (ed.). 2023. *Diccionario ortografico de l'aragonés* (versión preliminar). Zaragoza: Edicions Dichitals de l'Academia de l'Aragonés.

Dorgelès, Roland. 1919. *Croix de bois*. Paris: Éditions Albin Michel.

DNV = Acadèmia Valenciana de la Llengua. 2014. *Diccionari normatiu valencià*. València: Acadèmia Valenciana de la Llengua. www.avl.gva.es/lexicval/.

Erteschik-Shir, Nomi. 2019. Stage Topics and their Architecture. In V. Egerland, V. Molnar & S. Winkler (eds.), *The Architecture of topics*. Berlin: De Gruyter, Mouton.

Espinal, M. Teresa. 2009. Clitic incorporation and abstract semantic objects in idiomatic constructions. *Linguistics*, 47(6): 1221–1271.

— 2025. *Diccionari de sinònims de frases fetes*. Barcelona, València: Universitat Autònoma de Barcelona, Servei de Publicacions de la Universitat de València, Publicacions de l'Abadia de Montserrat. www.dsff.uab.cat.

Fabra, Pompeu. 1891. *Ensayo de gramática del catalan moderno*. Barcelona: L'Avenç.

— 1912. *Gramática de la lengua catalana*. Barcelona: L'Avenç.

— 1918. *Gramàtica catalana*. Barcelona: Institut d'Estudis Catalans.

— 1956. *Gramàtica catalana*. Barcelona: Teide.

Ferrando Francés, Antoni. 2024. El Llibre dels feits del rei en Jaume La versió canònica de la crònica jaumina. In A. Ferrando Francés, & V.J. Escartí Soriano, *Jaume I: Llibre dels feits*. València: Universitat de València, Acadèmia Valenciana de la Llengua.

Fischer, Susann & Esther Rinke. 2013. Explaining the variability of clitic doubling across Romance: a diachronic account. *Linguistische Berichte*, 236: 70–87.

Fontana, Josep M. 1993. Phrase Structure and the Syntax of Clitics in the History of Spanish. Ph.D. thesis, Philadelphia: University of Pennsylvania.

GNV = Acadèmia Valenciana de la Llengua. 2006. *Gramàtica normativa valenciana*. València: Acadèmia Valenciana de la Llengua. www.avl.gva.es/gnv/buscador.jsp?gramatica.

Greimas, Algirdas Julius. 1968. *Dictionnaire de l'ancien français jusqu'au milieu du XIV^e siècle*. Paris: Librairie Larousse.

Grevisse, Maurice & André Goosse. 2007. *Le Bon Usage. Grammaire française*. [14th ed.] Paris: De Boeck. Duculot.

Hale, Kenneth L. 1986. Notes on world view and semantic categories: Some Warlpiri examples. In P. Muysken & H. van Riemsdijk (eds.), *Features and Projections*, Dordrecht, Foris, 233-254.

Hale, Kenneth L., & Samuel Jay Keyser. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, MA: MIT Press.

Halle, Morris, & Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In K. Hale & S.J. Keyser (eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*, 111-176. Cambridge, MA: MIT Press.

Harley, Heidi. 2002. Possession and the Double Object Construction. *Linguistic Variation Yearbook*, 2: 29–68.

Harves, Stephanie, & Richard Kayne. 2012. Having ‘need’ and ‘needing’ have. *Linguistic Inquiry*, 43(1): 120–132.

Hopper, Paul J., & Elizabeth C. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

Iglesias, J. Antoni. 2008. Manuscrits i escriptura del “Llibre dels fets”. In *El “Llibre dels fets” digital (manuscrits, conquestes i llibres digitals)*. Recerca en acció: Fundació Institució Catalana de Suport a la Recerca. <https://www.recercaenaccio.cat/wp-content/uploads/documents/081202-manuscrits-i-escriptura-del-llibre-dels-fets-64053.pdf>.

Kim, Jaegwon. 1969. Events and their descriptions: Some considerations. In N. Rescher (ed.), *Essays in honor of Carl G. Hempel: A tribute on the occasion of his sixty-fifth birthday*, 198–215. Dordrecht: Reidel.

KIParla = Mauri, Caterina, Ballarè, Silvia, Gorla, Eugenio, Cerruti, Massimo, & Francesco Suriano. 2019. “KIParla corpus: a new resource for spoken Italian”. In: R. Bernardi, R. Navigli, & Giovanni Semeraro (eds.), *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it*. <https://kiparla.it/en/search/>.

Kratzer, Angelika. 1995. Stage-Level and Individual-Level Predicates. In G.N. Carlson, & F.J. Pelletier, *The Generic Book*, 176–223. Chicago: Chicago University Press.

— 1996. Severing the external argument from its verb. In J. Rooryck, & L. Zaring (eds.), *Phrase structure and the lexicon*, 109-137. Dordrecht: Springer.

Krifka, Manfred. 1989. Nominal reference, temporal constitution and quantification in event semantics. In R. Bartsch, J. van Benthem, & Peter van Emde Boas (eds.), *Semantics and contextual expression*, 75-115. Dordrecht: Foris.

Ledgeway, Adam. 2009. *Grammatica diacronica del napoletano*. Tübingen: Max Niemeyer Verlag.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago UP.

Maienborn, Claudia. 2001. On the position and interpretation of locative modifiers. *Natural Language Semantics* 9: 191–240.

— 2019. Events and states. In R. Truswell (ed.), *The Oxford handbook of event structure*, 50–89. Oxford: Oxford University Press.

Mangialavori Rasia, María Eugenia. 2021. Deconstructing the conative alternation in Romance. In C. Howe, P. Chamorro, T. Gupton, & Margaret Renwick (eds.), *Open romance linguistics* 1. Berlin: Language Science Press.

Mateu, Jaume. 2002. *Argument structure. Relational construal at the syntax-semantics interface*. Ph.D. thesis, Bellaterra: Universitat Autònoma de Barcelona.

Mistral, Frédéric. 1878 [1979]. *Lou tresor dóu Felibrige ou dictionnaire provençal-français* 2. Genève et Paris: Slatkine reprints.

Moignet, Gérard. 1973. *Grammaire de l'ancien français*. Paris: Editions Klincksieck.

Nagore, Francho. 2013. *Linguística diatópica del Alto Aragón*. Huesca: Consello d'a Fabla Aragonesa, Instituto de Estudios Altoaragoneses.

Nonell i Mas, Jaume. 1906. *Gramática de la llenga catalana*. Barcelona: Alvar Verdaguer.

Par, Anfós. 1923. *Sintaxi catalana segons los escrits en prosa de Bernat Metge (1398)*. Halle: Max Niemeyer Verlag; Beihefte zur Zeitschrift für Romanische Philologie.

Paradís, Anna (University of Oxford) (forthcoming).

Potts, Christopher. 2015. Presupposition and implicature. In S. Lappin, & C. Fox (eds.), *The Handbook of Contemporary Semantic Theory*, 168–202. New York: John Wiley and Sons.

Pylkkänen, Lina. 2002. *Introducing Arguments*. Ph.D. thesis, Cambridge: MIT.

— 2008. *Introducing Arguments*. Cambridge: MIT Press.

Ramchand, Gillian C. 2008. *Verb meaning and the lexicon. A first-phase syntax*. Cambridge: Cambridge University Press.

— 2017. The event domain. In R. D'Alessandro, I. Franco, & Ángel J. Gallego (eds.), *The Verbal Domain*. Oxford: Oxford University Press.

Rapoport, Tova. 2014. Central coincidence: The preposition *with*. *Prépositions et actualités. Faits de Langues: Revue de Linguistique* 44 : 159–173.

Real-Puigdollers, Cristina. 2013. Lexicalization by phase: The role of prepositions in argument structure and its cross-linguistic variation. Ph.D. thesis, Bellaterra: Universitat Autònoma de Barcelona.

— 2025. Low applicatives in Catalan: the case of *veure-hi*. *Borealis – An International Journal of Hispanic Linguistics* 14(1): 47–70.

Lipták, Anikó K., & Reintges, Chris H. 2006. Have=be+Prep(osition): New evidence for the preposition incorporation analysis of clausal possession. In M. Frascarelli (ed.), *Phases of Interpretation*, 112–128. Berlin, New York: De Gruyter Mouton.

Rigau, Gemma. 1997. Locative sentences and related constructions in Catalan: *ésser/haver* alternation. *Anuario del Seminario de Filología Vasca “Julio de Urquijo”* 40: 395–421. <https://doi.org/10.1387/asju.9572>.

— 2008. Els complements adjunts. In J. Solà, M.R. Lloret, J. Mascaró, & M. Pérez Saldanya (eds.), *Gramàtica del català contemporani*, 2045–2110. Barcelona: Empúries.

— 2017. Els pronoms febles adverbials, pronoms capritxosos?. Seminar script at the Centre de Lingüística Teòrica. Barcelona: Universitat Autònoma de Barcelona. <https://www.upf.edu/documents/8217493/127026344/17-18+lli%C3%A7o+UPF+traductor.pdf/080adeb8-9f55-6da8-d9eb-0672d2188ab5>.

— 2020. Presència i absència del clític locatiu “hi”. *Caplletra. Revista Internacional de Filologia*, 68: 99–116. <https://raco.cat/index.php/Caplletra/article/view/365846>. Russi.

Roberts, Ian & Anna Roussou. 2003. *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press.

Russi, Cinzia. 2008. *Italian Clitics. An Empirical Study*, Berlin, New York, Mouton de Gruyter.

Seki, Lucy. 2000. *Gramática do Kamaiurá: língua Tupi-Guarani do Alto Xingu*. São Paulo: Editora da Unicamp.

Solà, Joan. 1990. *Lingüística i normativa*. Barcelona: Empúries.

— 1994. *Sintaxi normativa: estat de la qüestió*. Barcelona: Empúries.

Sperber, Dan & Deidre Wilson. 1995. *Relevance. Communication and Cognition*. Oxford (UK), Cambridge (MA): Blackwell Publishers.

Stassen, Leon. 2013. Predicative possession. In M.S. Dryer, & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Plank Institute for Evolutionary Anthropology. wals.info/chapter/117.

TMAO = Billy, Dominique. 2017. *Trésor Manuscrit de l'Ancien Occitan*. CNRS, Université Toulouse Jean Jaurès, Association Internationale d'Études Occitanes. <http://tmao.aieo.org>.

Todolí, Júlia. 2008. Els pronoms. In J. Solà, M.R. Lloret, J. Mascaró, & M. Pérez Saldanya (eds.), *Gramàtica del català contemporani*, 2045–2110. Barcelona: Empúries.

Traugott, Elizabeth C. 1995. Subjectification in grammaticalization. In D. Stein & Wright, Susan (eds.), *Subjectivity and Subjectivisation: Linguistic Perspectives*, 31–54. Cambridge: Cambridge University Press.

Treccani = Istituto della Enciclopedia Italiana fondata da Giovanni Treccani. 2014. *Vocabolario Treccani*. www.treccani.it/vocabolario/.

Tsiakmakis, Evripidis & M. Teresa Espinal. 2022. Expletiveness in grammar and beyond, *Glossa: a journal of general linguistics* 7(1). <https://doi.org/10.16995/glossa.5807>.

van Gelderen, Elly. 2011. *The Linguistic Cycle: Language Change and the Language Faculty*. Oxford: Oxford University Press.

Villalba, Xavier. 2024. Expressive commitments: A normative approach to exclamations. *International Review of Pragmatics* 16(1): 87–108.

Sources of the examples

Ballarín, Angel. 1971. *Vocabulario de Benasque*. Zaragoza: Institución Fernando el Católico.

Brunenc, Uc. 2001 [13th c.]. Cuendas razos novelhas e plazens. In P. Gresti (ed.), *Il Trovatore Uc Brunenc, edizione critica con commento, glossario e rimario*. Tübingen: Niemeyer. [COM2].

Cànon d'Avicenna = Ibn Sīnā. *Quart del Cànon*. [anon. tr.]. In Cerdà Micó, Concepció. 2004. Lo Quart del Canon d'Avicenna: edició del text i estudi de les febres i l'orina. Ph.D. thesis, València: Universitat de València. [CICA].

Clams i Crims a la València medieval = M.A. Diéguez Seguí (ed.). 2002 [1279-1321]. *Clams i Crims en la València medieval segons el Llibre de Cort de Justícia (1279-1321)*. Alacant: Universitat d'Alacant. [CICA].

Costums de Tortosa. 1996 [1272]. J. Massip i Fonollosa (ed.). Barcelona: Fundació Noguera. [CICA].

de Bentadorn, Bernat. 1966 [13th c.]. “Ab joi mou lo vers e·l comens”. In M. Lazar (ed.), *Bernard de Ventadour, troubadour du XIIe siècle*. Paris: Klincksieck. [COM2].

de Commynes, Philippe. 1924 [1490-1505]. *Mémoires*. J. Calmette (ed.). Paris: Les Belles Lettres. <http://txm.ish-lyon.cnrs.fr/bfm/catalog/commyn3>. [BMF].

de Machaut, Guillaume. 1921 [1357]. Le confort d’ami. In *Oeuvres de Guillaume de Machaut*, v. 3. Paris: Librairie Ancienne Edouard Champion Éditeur. [DMF].

de Malla, Felip. *Memorial del pecador remut*. 1483. Girona: a despeses den Mateu Vendrell Mercader. bdh.bne.es/bnsearch/detalle/bdh0000174101. [CICA].

de Rouvroy, Louis. 1856 [1675-1755]. *Mémoires de Saint-Simon*. Paris: Librairie de L. Hachette et cie. [https://fr.wikisource.org/wiki/M%C3%A9moires_\(Saint-Simon\)](https://fr.wikisource.org/wiki/M%C3%A9moires_(Saint-Simon)). [Grevisse & Goose 2007].

DICPC = Oliveras i Duran, Sebastià (dir.). 2002. *Diccionari pràctic i complementari de la llengua catalana*. Barcelona: La Busca edicions.

Dorgelès, Roland. 1919. *Croix de bois*. Paris: Éditions Albin Michel.

Eiximenis, Francesc. 2005 [1379-1381; 1483]. *Dotzè del Crestià* (I), X. Renedo (ed.). Girona: Universitat, Diputació de Girona. [CICA].

Epistolari de Ferran I d’Antequera amb els infants d’Aragó i la reina Elionor (1413-1416). 2004 [1413-1416]. C. Rodríguez López (ed.). València: Universitat de València. [CICA].

Esteve i Puig, Pere. 2005 [17th c.]. *Escrits valencians*. València: Institució Alfons el Magnànim. [CIGCMod].

Gazull, Jaume [1497]. *Lo somni de Johan Johan*. València: Llop de la roca. https://weblioteca.uv.es/cgi/view.pl?source=uv_in_i17273353. [CICA].

La Germania. [1519-1525]. In V. Vallés Borràs. 2000. *La Germania*. València: Institució Alfons el Magnànim. [CICA].

Les Mystères provençaux du quinzième siècle. 1893 [15th c.]. A. Jeanroy, & H. Teulié (eds.). Toulouse: Imprimerie et librairie Édouard Privat. <https://gallica.bnf.fr/ark:/12148/bpt6k4149x.texteImage>. [COM2].

Llibre de Cort del Justícia de Cocentaina (1269-1290). 2009 [1269-1290]. J. Torró (ed.). València: Universitat de València. [CICA].

Llibre dels fets del rei en Jaume. 1991 [1343]. J. Bruguera (ed.). Barcelona: Barcino. [CICA].

Llull, Ramon. [1294]. *Taula general*. NARPAN.
https://narpan.net/ben/taulageneral_index.html. [CICA].

Lo Jutgamen General = Anon. 1971 [15th c.]. *Le Jugement dernier (lo Jutgamen General)*. *Drame provençal du XVe siècle*. M. Lazar (ed.). Paris: Klincksiek. [COM2].

Manfredi, Girolamo. 1474. *Liber de homine*. Bologna: Ugo Ruggeri, Donino Bertocchi. <https://archive.org/details/ita-bnc-in2-00001875-001/page/n2/mode/2up>.

— 2004 [1499]. *Quesits o perquens*. In A. Carré (ed.), *Quesits o perquens (Regiment de sanitat i tractat de fisiognomonía)*. Barcelona: Barcino. [CICA].

Manual de Consells de la ciutat de València. 2001 [1306-1326]. V. Anyó Garcia (ed.). València: Ajuntament de València. [CICA].

Martorell, Joanot. 2005 [1490]. *Tirant lo Blanc*. València: Tirant lo Blanch. [CICA].

Metge, Bernat. 2006 [1399]. *Lo somni*. S.M. Cingolani (ed.). Barcelona: Barcino. [CICA].

Muntaner, Ramon. *Crònica*. 2010 [1328]. J.A Aguilar (ed.). Ph.D. thesis, València: Universitat de València. [CICA].

Nosaltres sols = Editorial. 1932. “La Batalla per l’estatut. La traïció dels parlamentaris”. *Nosaltres sols!*, II, 62. Barcelona.
www.reeixida.cat/docs/ns/butlletins/NS062.pdf.

Olivier, Philippe. 2009. *Dictionnaire d’ancien occitan auvergnat. Mauriacois et Sanflorain (1340-1540)*. Tübingen: Max Niemeyer Verlag.

Pergamins, processos i cartes reials. Documentació dispersa valenciana del segle XIII. 2010 [1265-1399]. E. Guinot Rodríguez (ed.). València: Publicacions de la Universitat de València. [CICA].

Perrault, Charles. 2014. *Lo Capaironet roge*. Òussa: Letras d’Òc. [DICODÒC]

Procés criminal contra Antònia Marquès. 1998 [1374-1377]. J.A. Rabella (ed.). Barcelona: Curial. [CICA].

Quaresma de sant Vicent Ferrer predicada a València l’any 1413. 1927 [1413]. J. Sanchis Sivera (ed.). Barcelona: Institució Patxot. [CICA].

Questa del sant grasal. 2018 [ca. 1450-99]. J.Quingles i Fontcubeta (ed.). Palma: Edicions la Foradada – José J. de Olañeta, Editor. [CICA].

Roig, Jaume. 2006 [ca. 1460]. *Espill*. A. Carré (ed.). Barcelona: Quaderns Crema. [CICA].

Roís de Corella, Joan. 2018 [ca. 1348 i 1377]. *Lo cartoxà*. J.M. Furió, (ed.). València: Editorial Institució Alfons el Magnànim-Centre Valencià d'Estudis i d'Investigació.

The Towneley plays = Anon. 1996 [ca. 1550]. London, New York: Early English Text Society, Oxford University Press.

Un Llibre de Cort Reial mallorquí del segle XIV (1357-60). Introducció, transcripció i estudi lingüístic. 1984 [1357-1360]. J. Miralles i Monserrat (ed.). Palma: Institut d'Estudis Balearics, Editorial Moll. [CICA].