

# Between shared expectations and possibilities with discourse particles in European Portuguese – the case of *sempre*

João Azevedo

Universidade de Lisboa  
joaombailao@gmail.com

Jakob Maché

Centro de Linguística da Universidade de Lisboa  
jakob.mache@letras.ulisboa.pt

## 1 Introduction

This paper deals with the question of whether there are genuine ‘modal’ particles in Portuguese. It investigates candidates proposed in the literature such as *acaso*, *afinal*, *bem*, *cá*, *e*, *então*, *é que*, *já*, *lá*, *mas*, *não*, *se calhar*, confirmative *sempre* and *também*, as previously discussed by Franco (1990:175, 1998:147–149), Macário-Lopes (1998:8–10), Meisnitzer (2012:344–353) and peripheral or new candidates such as *embora* (cf. Pinto de Lima 1997) and *claro* investigated by Mendes and Lejeune (2022).

- (1) Vocês ontem sempre foram<sub>V</sub> ao  
you.PL yesterday after.all were at.the.M  
cinema?<sup>1</sup>  
cinema  
‘Did you go to the cinema after all?’
- (2) O banho de imersão é<sub>V</sub> sempre  
DET.M bath of immersion is always  
demorado.<sup>2</sup>  
take.time-PPP  
‘The full bath is always time consuming.’

To date, it is contested whether these items qualify as being of the syntactic category *particle* and to what extent they exhibit ‘modal’ semantics. This debate is largely due to terminological confusion inherited from the early pioneering work on German modal particles by Weydt (1969). For that reason the more general term *discourse marker* is used whenever the items do not neatly reflect the basic characteristics of the syntactic category particles or modal semantics.

The goal of this paper is to provide more clarity in the discussion based on corpus data and on recent

findings on discourse semantic. It concludes that at least the confirmative use of *sempre*, which will be the main focus of the study here, exhibits all the characteristics relevant for its German counterparts. Furthermore, regarding its semantics mirativity and the dimension of discourse and are much more relevant for its description than is the dimension modality in the narrow sense. In particular the particle *sempre* makes reference to previous *Discourse Commitments* and/or *Common Ground*, in the sense of Farkas and Bruce (2010:84–90). Finally, a definition of the syntactic category of discourse particles will be suggested which holds at least for Germanic and Romance languages.

## 2 Syntactic properties of *sempre*, *nem*, *lá*, *cá* and *afinal*

Adopting the more matured insights from decades of research on German modal particles discussed by Thurmair (1989:25–29, 36) and Coniglio (2008:14–16, 121–126, 148, 159), this paper shows that the term *particle* is essentially a syntactically motivated category. Specifically, it refers to a type of defective sentence adverbial, which has almost completely lost its capacity to be displaced within a clause and which lost its membership to any question domain, in the sense of Ginzburg (2012:122). And often also to a focus domain.<sup>3</sup>

As will be shown, confirmative *sempre* fulfils this criterion as it is bound to the preverbal position, but most of the other core candidates compete for the same slot (cf. Franco 1990:175, 1998:147, 150; Macário-Lopes 1998:7, Brito 2001:66 and Amaral

<sup>3</sup>However, there is a use of German *doch*, which attracts nuclear stress, similar as confirmative *sempre* in Portuguese. It remains to be investigated to which set of focus alternatives it belongs.

<sup>1</sup>CRPC-ORAL pfamev06.txt.

<sup>2</sup>CRPC-ORAL pf1202pu.txt.

and Del Prete 2014:137). We conducted two independent corpus studies, which are not published yet and which are based on data from the *DiLeB* corpus (*Discurso Informal de Lisboa e Braga*) and the *CRPC* corpus (*Reference Corpus of Contemporary Portuguese*).<sup>4</sup>

The frequencies are represented in Tables 1 and 2. Note that it is a major characteristic of discourse particles that they are only one use of polyfunctional or heteronym lexemes, which are often extremely difficult to distinguish (Thurmair 1989:21, Helbig 1994, Coniglio 2008:8–9). In most cases, there are also less grammaticalised, more lexical adverbial uses. For instance Portuguese *sempre* also has a dominant use as a temporal adverb with the meaning of ‘always’ that universally quantifies over time intervals.

The figures are just counts of the lexemes without differentiating the particle uses. However, the discourse-oriented uses of these adverbs are rare, and almost exclusively in the preverbal position. The situation with *afinal* is less clear, because it always denotes discourse meaning but it at times occurs postverbally and clause initially and unlike *sempre* it can occur with different placements of the nuclear stress (cf. Amaral and Del Prete 2014:141 for a similar observation for its Italian counterpart *alla fine*). In the *DiLeB* corpus, it has a strong preference for the preverbal position, in the *CRPC* corpus, however it is attested mostly in clause initial position. Likewise, the discours-oriented uses of *nem* have a strong preference for the preverbal position.

Being limited to the preverbal position, confirmative *sempre* behaves exactly like other recognised modal particles in other Romance such as Italian *mai* and particles in Germanic languages such as German and Scandinavian (cf. Coniglio 2008:14–16, 95–101, Coniglio 2023:11–15). In other words, particles defined by their placement restrictions to some posi-

<sup>4</sup>The *DiLeB*-corpus (*Discurso Informal de Lisboa e Braga*) is an online corpus of informal conversations with speakers from Lisbon and Braga in sociolinguistic interviews from the last decade of the 20<sup>th</sup> century.

<http://teitok.clul.ul.pt/dileb/index.php?action=home>

The *CRPC* corpus (*CRPC-ORAL (Reference Corpus of Contemporary Portuguese)*) is another online corpus composed mostly of spoken European Portuguese from various contexts and periods of recent time.

<http://teitok.clul.ul.pt/crpcoral/index.php?action=home>

position	<i>nem</i>	<i>sempre</i>	<i>afinal</i>	<i>lá</i>	<i>cá</i>
preverbal	181	177	13	505	132
postverbal	99	503	2	1018	259
other pos.	283	109	0	345	115
total	563	789	15	1868	506

Table 1: The placement of discourse modifiers and their polyfunctional variants – *CRPC* corpus

position	<i>nem</i>	<i>sempre</i>	<i>afinal</i>	<i>lá</i>	<i>cá</i>
preverbal	395	288	7	1232	133
postverbal	152	1009	5	2358	378
other pos.	523	297	25	867	334
total	1070	1594	37	4457	845

Table 2: The placement of discourse modifiers and their polyfunctional uses – *DiLeB* corpus

tion at the left edge of the extended VP or TP, which corresponds to the the preverbal position in Romance languages, and some position in the middle field in German.<sup>5</sup>

It is quite plausible to assume that the extremely limited freedom to move these particles can be derived from more general requirements. As observed by Hentschel (1983:48, 1986:210–213, 232–238) Thurmair (1989:25–37) and Coniglio (2008:102–108), the dimensions of givenness and definiteness play an essential role in the placement of discourse particles in German. These authors argue that discourse particles mark the boundary between the theme (given information) and rheme (discourse new information). Their observations suggest that the nuclear stress has to follow the particles, and that it also might be the focus, rather than the rheme, which follows the particle.

Despite Franco’s (1998:144, 150) observes that

<sup>5</sup>However, there are no discrete boundaries between discourse particles and sentential adverbs as regards to their semantics. Confirmative *sempre* in Italian is fairly common in postverbal position, which is only possible for adverbs under the definition defended here. But at the same time, Italian *sempre* lacks past related interpretations in as demonstrated by Amaral and Del Prete (2014:135, 137, 140–149), which is a clear indicator that it is grammaticalised to a lesser degree. In a similar vein, Portuguese *afinal* and its Italian counterpart *alla fine* occur in positions typical for adverbs but nevertheless these itmes display a meaning related to *sempre*, indicating an epistemic change or conflict in the discourse.

discourse particles in Portuguese do not mark the left boundary of the constituent of the rheme, the Portuguese confirmative particle *sempre* is also sensitive to focus. Amaral and Del Prete (2014:139–140, 148–149) assume that the confirmative use of *sempre* requires the entire prejacent proposition to be destressed, while *sempre* itself attracts the nuclear stress. As such it behaves semantically very similar to the stressed version of the German particle *doch* (cf. Egg and Zimmermann 2012), as already noticed by Franco (1998:153).

### 3 Modal or discourse oriented – the semantics of *sempre*

Thurmair (1989:3) observed that previous studies on particles used the term ‘modal’ loosely to indicate that these markers convey extra-propositional meaning, following Palmer’s (1986:1) very vague definition of modality, which posits that modal modifiers encompass any type of modifier that take scope over the proposition. Consequently, this term does not significantly contribute much in the characterisation of these elements. In subsequent research, Portner (2009:1) proposed a much clearer and accurate definition according to which modality makes statements about situations in non-actual possibilities/possible worlds. He draws much of his inspiration from earlier work by Kratzer (1978, 1981), who focusses only on selected modal verbs in German and English without providing an explicit definition of modality as a phenomenon itself. However, among the 20–35 markers considered as modal particles by Thurmair (1989:49) and Durell (2011) only a few, such as the epistemic particles *wohl* refer to non-actual possibilities. Most of these elements instead refer to shared or individual beliefs, convictions or statements by discourse participant or third participants, some of them but not all also refer to events in non-actual worlds.

The dimension of discourse appears to be much more relevant for many of the items under discussion since they reference to propositions to which at least one of the speech participant publicly committed too. Macário-Lopes (1998:8–9) observed that the confirmative particle *sempre* expresses expectations and doubts from the speaker’s perspective regarding the truth of the prejacent proposition *p*. Amaral and Del Prete (2014:135–140, 2016:1135–1137, 2020:5–7) propose a more specific and detailed description of the semantic contribution of confirmative *sempre*.

They argue that it is “only felicitous in a context where the truth of the prejacent is presupposed to have been under discussion by the interlocutors” (p. 140).

In their analysis, Amaral and Del Prete (2014:149–150) take the particle *sempre* to be an epistemic modal operator that takes a proposition *p* and returns the confirmation of the truth of that proposition, and referencing three different times: a point  $t_1$  prior to utterance time when the prejacent  $p_4$  was considered true in all the best epistemic worlds, a subsequent point  $t$  when it became possible that *p* could be false, and at utterance time when *p* is confirmed to be true. Crucially, they align with Kratzer’s (1978, 1981) view that epistemic modal operators are evaluated with respect to collective knowledge rather than individual knowledge.

Amaral and Del Prete’s (2014:149–150) analysis captures many of the relevant aspects of *sempre*. But it has two essential shortcomings. Firstly, it does not distinguish between propositions to which both speech participants commit (hence part of the common ground) and propositions which are merely known to the speech participants without that both of them committed to their truth. The most common case are propositions to which only one speech participant committed publicly by asserting it, but whose truth is contested by the other party. This conflict can be resolved assuming that there is another resource in the discourse representation besides the common ground, which represents shared beliefs, namely discourse commitments, in the sense of Farkas and Bruce (2010) that is, propositions to which only one of the party is committed, too.

The second shortcoming concerns the conception of epistemic modality. Lasersohn (2005:277), Stephenson (2007:489) and Maché (2013:405–422) observed that epistemic modal operators are not evaluated with respect to what is “generally known in the world”, as suggested in Kratzer’s early work but their interpretation is always dependent on a specific epistemic judge.

Amaral and Del Prete (2014:149–150) face challenges in explaining the most common occurrences of *sempre*, particularly when there is a disagreement at  $t_2$  between the speaker and addressee or third party about the truth of the proposition *p*. This is illustrated in example (3). The holder of the expectation *p* can be explicitly encoded by phrases like ‘you were right’ and as shown below, speaker and addressee can have different convictions:

- (3) Tinha/Tinhas/A minha  
 have.IMP.1S/have.IMP.2S/DET.F my.F  
 vezinha tinha razão, o  
 neighbour have.IMP.23 right DET.M  
 CHEGA sempre ganhou mais de  
 Chega sempre win.PST.3S more than  
 15%.  
 15%  
 ‘I was/You were/My neighbour was right, the  
 CHEGA party won more than 15% after all.’

The proposition ‘CHEGA wins more than 15%’ in the example above was never part of the *Common Ground*, because it was never shared belief. A similar reasoning applies to the belief at  $t_2$  that  $p$  could be false, which can be demonstrated with utterances of the type *T’as a ver?* ‘Do you see now?’. In other words, the roles of the expectation holder and the questioner can be assigned to speaker and addressee or vice versa.

## 4 Analysis

The analysis here is implemented in HPSG<sub>TTR</sub>/KoS (based on *Type Theory with Records*) as being suggested by Ginzburg (2012) and Lücking, Ginzburg, and Cooper (2021), which offers significant advantages for representing discourse relations.<sup>6</sup>

Confirmative *sempre* is considered a defective sentential adverb with the denotation of a propositional modifier of the type  $\langle t, t \rangle$ , which takes a proposition and returns the same proposition and adds the requirement that two specific previous moves must have had occurred for its truth conditions to be met: a public commitment to the truth of  $p$  by some speech participation and the concession that  $p$  might be false at some later move. The semantic contribution is summarised as follows. Confirmative *sempre* makes reference to three times:  $t_1$ , which precedes  $t_2$ , which in turn precedes  $t_{utt}$ . There is an epistemic attitude holder  $x$  who publicly committed to the truth of  $p$  or that it would become true at  $t_1$ , and at some later moment  $t_2$  the retracted the public commitment and concedes the possibility that  $p$  could be false. This

<sup>6</sup>An anonymous reviewer wondered why the analysis here was not formulated within the framework of *Lexical Resource Semantics*. The decision in favour of HPSG<sub>TTR</sub> is motivated by the fact that it is much more appropriate to model the semantics of dialogues, including for instance the MOVE feature, which tracks all the utterances of all speech parties that were previously made.

is done by means of a modal operator which is anchored to some attitude holder or modal judge  $x$ , in the sense of Stephenson (2007:501). Crucially, both the expectation and the concession that  $p$  might be false are moves in the previous dialogue. Eventually the speaker asserts  $p$ . Note that the speaker can be identical to  $x$  or not.

This analysis allows for a similar treatment with question bias with negative polar questions with low negation reading, as analysed by Sudo (2013:276–284). Speakers who utter this type of question express that they originally believed that  $p$  was true (epistemic bias) but shortly before utterance time they were confronted with compelling evidence suggesting that  $p$  is false (evidential bias). It is assumed here that the situation is the parallel with *sempre*. Instead of giving a detailed account of the type of evidence made the attitude holder change their mind, the change of epistemic state is modelled by the introduction of an epistemic possibility at  $t_2$ , without specifying what evidence made that change happen.

Implemented in HPSG<sub>TTR</sub>, the lexicon entry for confirmative *sempre* is illustrated in Figure 1. It is assumed here that there was a previous move  $m_1$ , in which one of the speech participants  $x$  which can be the speaker or the addressee committed publicly to the truth of the proposition  $p$  by asserting it. At some subsequent move  $m_2$  the same speech participant retracts their public commitment and concedes the possibility that  $\neg p$  (in case  $y = x$ ). But  $y$  can also be instantiated by another speech participant, then the retraction of the commitment to the truth  $p$  is not necessary. As the expectation that  $p$  is or becomes true is modelled as discourse commitment by means of some move made earlier in the dialogue, the analysis here is compatible with  $p$  being part of the *Common Ground* in some scenarios.

The move  $m_2$  contains a possibility operator following Cooper’s (2023:247) translation of Kratzer’s (1981) concept of modal operators into TTR. In the present study it is a predicate of the arity  $\langle \text{Ind}, \text{Type}, \text{Type}, \text{Type} \rangle$ , taking as arguments: an epistemic judge (here modelled as *Individual*), a proposition (here modelled as *Type*), an epistemic modal base  $B$  (here modelled as *Type*) and a stereotypical ordering source of ideals  $I$  (here modelled as a *Type*, too).

Summing up the semantics, in many ways confirmative *sempre* resembles stressed *doch* in German which references to a negated proposition that was earlier present in the common ground, as illustrated by Karagjosova (2009) Egg and Zimmermann

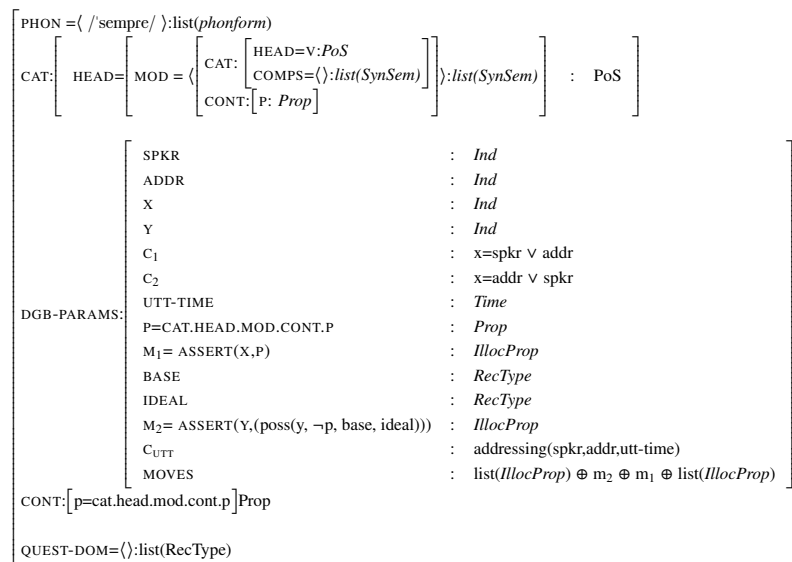


Figure 1: Lexicon entry for the confirmative discourse particle *sempré*

(2012:227–228) and Döring (2016). Returning to the question, whether European Portuguese has items that deserve to be considered as discourse particles of the Germanic type, the answer is clearly yes.

Amaral, Patrícia and Fabio Del Prete (2014). “On truth persistence. A comparison between European Portuguese and Italian in relation to *sempré*”. In: *Variation within and across Romance languages. Selected papers from the 41<sup>st</sup> linguistic symposium on Romance languages*. Ed. by Marie-Hélène Côté and Eric Mathieu. Current Issues in Linguistic Theory 333. Amsterdam: John Benjamins Publishing Company, pp. 135–154.

- (Jan. 2016). “On truth unpersistence: At the crossroads of epistemic modality and discourse”. In: *Natural Language & Linguistic Theory* 34.4, pp. 1135–1165. ISSN: 1573-0859. URL: <http://dx.doi.org/10.1007/s11049-015-9325-5>.
- (Oct. 2020). “Predicting the end: Epistemic change in Romance”. In: *Semantics and Pragmatics* 13.11, pp. 1–50. ISSN: 1937-8912. URL: <http://dx.doi.org/10.3765/sp.13.11>.

Brito, Ana Maria (2001). “Clause structure, subject positions and verb movement. About the position of *sempré* in European Portuguese and Brazilian Portuguese”. In: *Romance languages and linguistic theory 1999*. Ed. by Yves d’Hulst, Johan Rooryck, and Jan Schroten. Amsterdam: John Benjamins Publishing Company, pp. 63–85.

Coniglio, Marco (2008). “Die Syntax der deutschen Modalpartikeln: ihre Distribution und Lizenzierung in Haupt und Nebensätzen”. PhD thesis. Università Ca’ Foscari Venezia.

- (2023). *On a syntactic definition of modal particles. Synchronic and diachronic reflections*. Presentation given at the Workshop *Modal particles in Romance languages* at the Universidade de Lisboa 1<sup>st</sup>–2<sup>nd</sup> June 2023.

Cooper, Robin (2023). *From perception to communication – A type theory for action and meaning*. Oxford Studies in Semantics and Pragmatics 16. Oxford University Press.

Döring, Sophia (2016). “Modal particles, discourse structure and common ground management. Theoretical and empirical aspects”. PhD thesis. Berlin: Humboldt Universität zu Berlin.

Durell, Martin (2011). *Hammer’s GERMAN Grammar and Usage*. 5th ed. London: Routledge.

Egg, Markus and Malte Zimmermann (2012). “Stressed out! Accented discourse particles: The case of *doch*”. In: *Proceedings of Sinn und Bedeutung 16*. Ed. by Ana Aguilar Guevara, Anna Chernilovskaya, and Rick Nouwen. Konstanz: Open Journal Systems, pp. 225–238. URL: <https://ojs.uni-konstanz.de/sub/index.php/sub/article/view/422/355>.

Farkas, D. F. and K. B. Bruce (Sept. 2010). “On Reacting to Assertions and Polar Questions”. In: *Journal of Semantics* 27.1, pp. 81–118. URL: <https://doi.org/10.1093%2Fjose%2F27fp010>.

Franco, António (1990). “Partículas Modais do Português”. In: *Revista da Faculdade de Letras : Línguas e Literaturas* 7, pp. 176–196.

- (1998). “Partículas Modais da Língua Portuguesa: Relances Contrastivos com as Partículas Alemãs”. In: *Revista da Faculdade de Letras : Línguas e Literaturas* 5.1, pp. 137–156.

Ginzburg, Jonathan (2012). *The interactive stance: meaning for conversation*. Oxford University Press.

Helbig, Gerhard (1994). *Lexikon deutscher Partikel*. Leipzig: Langenscheidt.

Hentschel, Elke (1983). “Partikeln und Wortstellung”. In: *Partikeln und Interaktion*. Ed. by Harald Weydt. Reihe Germanistische Linguistik 44. Tübingen: Max Niemeyer, pp. 46–53.

- (1986). *Funktion und Geschichte deutscher Partikeln*. Ja, doch, halt und eben. Reihe Germanistische Linguistik 63. Tübingen: Max Niemeyer.

Karajosova, Elena (2009). “A unified DRT-based account of accented and unaccented middle field *doch*”. In: *Sprache & Datenverarbeitung* 33.1–2, pp. 77–93.

Kratzer, Angelika (1978). *Semantik der Rede. Kontexttheorie – Modalwörter – Konditionalsätze*. Scriptor.

- (1981). “The Notional Category of Modality”. In: *Words, Worlds and Contexts. New approaches in World Semantics*.

- Ed. by Hans Jürgen Eikmeyer and Hannes Rieser. Berlin: Mouton de Gruyter, pp. 38–74.
- Lasersohn, Peter (2005). “Context dependence, disagreement and predicates of personal taste”. In: *Linguistics and Philosophy* 28, pp. 643–686.
- Lücking, Andy, Jonathan Ginzburg, and Robin Cooper (2021). “Grammar in dialogue”. en. In: *Head-Driven Phrase Structure Grammar: The handbook*. Ed. by Stefan Müller et al. Berlin: Language Science Press, pp. 1155–1199. URL: <https://zenodo.org/record/5599870>.
- Macário-Lopes, Ana Cristina (1998). “Contribuição para o estudo dos valores discursivos de sempre”. In: *Actas do XIII Encontro Nacional da APL*. Ed. by M. A. Mota and R. Marquilhas. Lisboa: Colibri, pp. 3–14.
- Maché, Jakob (2013). “On black magic. How epistemic modifiers emerge”. PhD thesis. Freie Universität Berlin.
- Meisnitzer, Benjamin (2012). “Modality in the Romance languages: Modal verbs and modal particles”. In: *Theory of Mind Elements across Languages*. Ed. by Werner Abraham and Elisabeth Leiss. Trends in Linguistics 243. Berlin: de Gruyter, pp. 335–360.
- Mendes, Amália and Pierre Lejeune (2022). *Marcadores discursivos e marcadores modais: uma análise contrastiva de claro e seus equivalentes funcionais em francês*. Talk given at Associação Portuguesa de Linguística.
- Palmer, Frank Robert (1986). *Mood and Modality*. Cambridge text books in linguistics. Cambridge: Cambridge University Press.
- Pinto de Lima, José (1997). “Caminhos semânticos-pragmáticos da gramaticalização: O case de *embora*.” In: *Sentido que a Vida Faz. Estudos para Óscar Lopes*. Ed. by Ana Maria Brito et al. Porto: Campo das Letras, pp. 643–655.
- Portner, Paul (2009). *Modality*. Oxford: Oxford University Press.
- Stephenson, Tamina (2007). “Judge Dependence, Epistemic Modals and Predicates of Personal Taste”. In: *Linguistics and Philosophy* 30.4, pp. 487–525.
- Sudo, Yasutada (2013). “Biased polar questions in English and Japanese”. In: *Beyond expressives: Exploration in use-conditional meaning*. Ed. by Daniel Gutzmann and Hans-Martin Gärtner. Leiden: Brill, pp. 275–295.
- Thurmair, Maria (1989). *Modalpartikeln und ihre Kombinationen*. Linguistische Arbeiten 223. Max Niemeyer.
- Weydt, Harald (1969). *Abtönungspartikel. Die deutschen Modalwörter und ihre französischen Entsprechungen*. Bad Homburg: Gehlen.