

CogniTextes

Revue de l'Association française de linguistique cognitive

Volume 25 | 2024

Constructions and contexts

When constructional choice is a matter of context: *sembrare*-constructions across a continuum of text genres

FLAVIO PISCIOTTA

<https://doi.org/10.4000/12rcy>

Résumés

English Français

SEEM-verbs constructions, such as those with Italian *sembrare* 'seem', have been widely studied, with much of the research investigating the alternation between impersonal constructions with finite complementation, and personal ("raising") constructions governing an infinitival clause. However, most of the studies on the matter seem to disregard the role of the extra-linguistic context of production in the choice between said alternants, even though in various languages there seems to be a strong variability between written and spoken situations (Davies 1997a, b; Cornillie 2007; Pisciotta 2023). The aim of this study is thus to evaluate the relevance of extra-linguistic context (operationalized as text genre), as opposed to linguistic factors, in the choice between *mi sembra che* 'it seems to me that', *sembra che* 'it seems that' and *sembrare*+infinitive in Italian. We do so by integrating a Construction Grammar perspective with studies on register variation and on orality and scriptuality, adopting a point of view that decomposes written and spoken language in a continuum of text genres (Voghera 2017). A multifactorial analysis is carried out on 405 total occurrences of (*mi*) *sembra che* and *sembrare*+infinitive constructions, extracted from written and spoken Italian corpora, by fitting a Conditional Inference Tree. The results show that, at least for *mi sembra che* and *sembrare*+infinitive, the heaviest role in determining the choice of the constructions is played by the degree of dialogicity and planning of text genres, which are split by the model in three groups: conversational speech, an intermediate group consisting of fiction prose and monological speech, and non-fiction prose. Furthermore, data reveal that a continuum-based perspective on text genres allows us to interpret and to draw generalizations on the behaviour of *sembrare*-constructions. These results encourage us to take into account and to encode text genres in the description of constructions, even "traditional" and less idiosyncratic ones.

Les constructions avec des verbes de type *sembler*, comme celles avec l'italien *sembrare*, ont été largement étudiées. Une grande partie des recherches se concentre sur l'alternance entre les constructions impersonnelles avec une complémentation finie et les constructions personnelles (« à montée ») régissant une proposition infinitive. Cependant, la plupart des études sur le sujet semblent négliger le rôle du contexte extra-linguistique dans le choix entre ces alternants, bien



que dans diverses langues, il semble y avoir une forte variabilité entre les situations écrites et parlées (Davies 1997a, b; Cornillie 2007; Pisciotto 2023). L'objectif de cette étude est donc d'évaluer la pertinence du contexte extra-linguistique (opérationnalisé comme genre textuel), par opposition aux facteurs linguistiques, dans le choix entre *mi sembra che* ('il me semble que'), *sembra che* ('il semble que') et *sembrare*+infinitif en italien. Nous intégrons pour cela la perspective des Grammaires des Constructions avec des études sur la variation de registre et sur l'oralité et la scripturalité, en adoptant un point de vue qui décompose la langue écrite et parlée dans un continuum de genres textuels (Voghera 2017). Une analyse multifactorielle est réalisée sur un total de 405 occurrences des constructions (*mi*) *sembra che* et *sembrare*+infinitif, extraites de corpus italiens écrits et parlés, à l'aide d'un arbre de décision conditionnelle. Les résultats montrent que, du moins pour *mi sembra che* et *sembrare*+infinitif, le rôle le plus déterminant dans le choix des constructions est joué par le degré de dialogicité et de planification des genres textuels, que le modèle divise en trois groupes : discours conversationnel, un groupe intermédiaire constitué de la prose fictionnelle et du discours monologique, et la prose non fictionnelle. De plus, les données révèlent qu'une approche basée sur un continuum des genres textuels nous permet d'interpréter et de généraliser le comportement des constructions avec *sembrare*. Ces résultats nous encouragent à prendre en compte et à encoder les genres textuels dans la description des constructions, même celles dites « traditionnelles » et moins idiosyncratiques.

Entrées d'index

Mots-clés : genre textuel, grammaire des constructions, verbes de type sembler, choix constructionnel, alternance syntaxique

Keywords: text genres, construction grammar, SEEM-verbs, constructional choice, syntactic alternations

Texte intégral

1. Introduction

- 1 The constructions with SEEM-verbs (such as *sembrare* 'seem') have received wide attention in syntax for their peculiar behaviour (e.g., Postal 1970; Kayne 1981; Moro 1992; Torrego 1996; Rooryck 1997). Primarily, research has focused on the constructions with a complement clause, both finite (1a) and non-finite (1b)¹.

- (1) a. (Mi) Sembra che Gianni dorma
 to-me seem.IND.PRES.3SG COMP Gianni sleep.SBJV.PRES.3SG
 'It seems (to me) that Gianni is sleeping'
- b. Gianni (mi) sembra dormire
 Gianni to-me seem.IND.PRES.3SG sleep.INF
 'Gianni seems (to me) to be sleeping'

- 2 Many studies have considered the two constructions synonymous: scholars working in formal frameworks claim that the two sentences (1a-b) are the result of different transformations applied to the same deep structure (Rosenbaum 1967; Postal 1974; Polinsky 2013). On the contrary, the treatments of the (overt or omitted) experiencer *a me/mi* 'to me' turn out to be mixed: according to some analyses, the experiencer, although not expressed, is always present in the logical structure of the predicate *sembrare* 'seem' (Montague 1970), and some consider it as a default argument (e.g., Musi 2016), while for others the experiencer is not an argument of the predicate 'seem', but an external argument adjoined to the clause (Torrego 1996).
- 3 Functional and constructionist studies dealing with these constructions have shown that, despite the truth-conditional equivalence of *sembra che* (-like) and *sembrare*+infinitive (-like) constructions, functional differences motivate the formal

differences between them. Some studies have focused on their evidential and epistemic values, sometimes also addressing the presence or absence of the experiencer (Cornillie 2007; Gisborne & Holmes 2007; Aijmer 2009; Musi 2015; Mortelmans 2017). Moreover, in Pisciotto (2023), constraints regarding the NPs filling the subject slot were found to favour the use of either *sembra che* or *sembrare*+infinitive in Italian, similarly to their counterparts in other languages (Davies 1997a; 1997b; Grimm 2010). The analysis, carried out by means of monofactorial tests, revealed that *sembrare*+infinitive tend to select heavier subjects (2) than (*mi*) *sembra che* (that shows instead a higher rate of zero subjects), but also that it is preferred in the context of subject relativization (3a-3b). Moreover, it almost never appears with non-3rd person subjects, while the use of 1st and 2nd person subjects is less constrained with (*mi*) *sembra che* (4a-4b).

- (2) a. **La maggioranza degli utenti di Internet** sembra aver scoperto ISDN [...] solo di recente.

'The majority of Internet users seem to have discovered ISDN [...] only recently.'
(CORIS, MISC)

- (3) a. [...] un indizio // **che sembra supportare** l'idea di meccanismi comuni alle diverse forme neuropatologiche

'a clue **that seems to support** the 'idea of mechanisms common to different neuropathological forms' (KIParla, BOD1004)

- b. ?un indizio **che sembra che supporti** l'idea di meccanismi comuni alle diverse forme neuropatologiche

lit. 'a clue **that seems that supports** the 'idea of mechanisms common to different neuropathological forms'

- (4) a. sembra che [tu] mi stai intervistan[do]

'It seems that you are interviewing me' (KIParla, BOA3003)

- b. ?[Tu] Sembri starmi intervistando.

'You seem to be interviewing me'

- 4 The analysis of these constraints made it possible to uncover some functional idiosyncrasies of these constructions. Nonetheless, another possible factor determining the choice between these constructions in Italian seems to be the extra-linguistic context: there is a strong discrepancy between their frequencies in written and spoken situations (Pisciotto 2023). Since extra-linguistic context is often disregarded in the study of syntactic alternations, we want to assess its relevance in the case of *sembrare*-constructions. However, we hypothesize that the notions of written and spoken language are not specific enough to adequately explain variation in the frequency of choice between the constructions with *sembrare*. Instead, we suggest to "decompose" written and spoken contexts in order to carry out a more fine-grained analysis, by taking into account the specific communicative situations (i.e., text genres²) with their own properties (Biber 1986a; 1986b; Chafe & Tannen 1987). Hence, from a methodological point of view, we will integrate Construction Grammar (henceforth CxG) with tools and reflections coming from corpus linguistics studies on register variation and from the Italian tradition of studies on orality and scriptuality. Theoretically, we will investigate the role and the relevance of extra-linguistic context in the description of "normal syntax" (Hilpert 2019: 68), i.e., traditionally studied syntactic constructions. We will do so by analyzing a sample of (*mi*) *sembra che* and *sembrare*+infinitive constructions, extracted from both written and spoken Italian corpora. A Conditional Inference Tree will be fit including both linguistic (e.g., animacy of the subject, referent accessibility of the subject etc.) and extra-linguistic variables (i.e., written or spoken corpus, text genre), in order to find which factors influence most the choice between *sembrare*-constructions. More generally, this will also help us

understand the role of extra-linguistic factors in the study of alternations: that is, if No Synonymy between constructions (Goldberg 1995) can be granted not only via linguistic constraints, but also on contextual grounds. The possible “primacy” of contextual factors in constructional choice would indeed pose the question of adequately representing them in speakers’ linguistic knowledge.

5 The paper is organized as follows: in Section 2 we will give an overview of previous studies on the influence of context on SEEM-verbs constructions; in Section 3 we will discuss the notions of writing and speech and propose text genres as the ideal environment to study the relationship between constructions and extra-linguistic context, focusing in particular on continuum-based approaches to text genres; in Section 4 we will describe the dataset and the statistical methodologies; in Section 5 we will carry out an analysis to investigate the amount of influence of contextual factors, as compared to linguistic ones, in predicting the choice of our constructions; in Section 6 we will apply the results of the analysis to our constructions; finally in Section 7 we will discuss the implications of our results in a wider fashion.

2. Context and SEEM-verbs constructions: opposing written and spoken situations?

6 Even though studies on SEEM-verbs in Italian are not lacking, research has not usually focused in a systematic way on the influence of extra-linguistic context on the constructions in which these verbs appear. While this partly depends on the fact that hardly any study on the topic has been carried out taking into account spoken data, this could be also due to the fact that the majority of these studies have investigated SEEM-verb constructions as argumentative strategies, thus limiting the analyses to strongly argumentative text genres, such as opinion articles and reviews (Miecznikowski & Musi 2015; Musi 2015; Miecznikowski 2020). An exception can be found in Kratschmer (2006), that makes some observations concerning stylistic factors. The author’s findings rely on speakers’ judgements on corpus occurrences of *sembra che* or *sembrare*+infinitive and on their paraphrases with the other construction. The results show that the use of *sembrare*+infinitive is perceived as more sophisticated and “literary”, whereas the use of *sembra che* belongs to a more neutral style.

7 Apart from stylistic aspects (whose definition as “literary” or “neutral” is quite ambiguous) the only hint at a possible difference between contexts of use can be found in a previous study by the same author (Kratschmer 2005). In this case, the scholar looks at the differences in the distribution of the two constructions in two different novels and links the use of *sembrare*+infinitive to third-person narrations and the use of *sembra che* to first-person narrations that resemble direct oral discourse. However, these claims are not substantiated by a balanced or representative sample, and thus can only be seen as hypotheses.

8 Some data on the distribution of the constructions with *sembrare* in written and spoken Italian can be found in Pisciotta (2023). In this study, three constructions are taken into account: *sembra che*, *sembrare*+infinitive and *mi sembra che* ‘it seems to me that’ (the impersonal construction with the experiencer). While the study focuses on semantic and syntactic constraints, a relevant aspect is that the frequency of the three constructions shows a strong discrepancy between the written and spoken data³ (Table 1).

	<i>sembra che</i>	<i>mi sembra che</i>	<i>sembrare</i> +infinitive ⁴	Total
Written	61 (28,24%)	24 (11,11%)	131 (60,65%)	216 (100%)
Spoken	63 (33,33%)	100 (52,91%)	26 (13,76%)	189 (100%)

Total	124	124	157	405
--------------	-----	-----	-----	-----

Table 1. Frequency distribution of sembrare-constructions in written and spoken Italian (adapted from Pisciotta 2023: 104)

9 As we can see in Table 1, *sembrare*+infinitive is the most frequently used construction in writing, while it appears very rarely in the speech corpus. On the contrary, *mi sembra che* turns out to be the most used construction in the speech corpus and the least used in written one. Finally, *sembra che* seems to be almost neutral with respect to this distinction.

10 This distribution poses questions on which properties of written and spoken modes determine the use of *sembra che*, *mi sembra che* and *sembrare*+infinitive. In order to find a preliminary explanation of the influence of context on SEEM-verbs constructions, we can look at studies on other languages that display a distribution of these constructions similar to Italian, namely Spanish (Davies 1997b; Cornillie 2007) and Portuguese (Davies 1997a)⁵.

11 Davies, working on both languages (1997a; 1997b), hypothesises that this discrepancy is related to differences in language production between writing and speech. Referring to several studies, he proposes that in speech on-line production favours less integrated and shorter sentences connected by subordinating conjunctions (such as *parece que* 'it seems that'), rather than more integrated, complex and coherent sentences (*parecer*+infinitive 'seem+infinitive'). While this seems a reasonable hypothesis, it accounts only partially for our data, because it explains the distribution of the infinitive construction, but it does not make a distinction between *parece que* with and without the experiencer.

12 A different analysis comes from Cornillie's (2007) study on the constructions with *parecer* in Spanish. Differently from Davies, Cornillie draws a distinction between the impersonal construction with (*me parece que* 'it seems to me that') and without (*parece que*) an overt experiencer. This distinction proves to be fruitful from a functional point of view and provides a frequency distribution similar to the one in Table 1: in the written corpus, the most used construction is the infinitive one, while in the spoken corpus the most frequent one is the impersonal construction with overt experiencer. However, a difference is that the impersonal constructions without experiencer also show a more pronounced increase in frequency in the spoken corpus. Cornillie hypothesizes that this distribution is due to functional differences between constructions that are thus favoured or disfavoured by the nature of the speaker-listener interaction (2007: 17). In a detailed analysis of the evidential values of constructions with *parecer*, he notes that in *parecer*+infinitive the speaker's judgment is always subjective (i.e., it is construed only by the speaker), whereas in *parece que* a more intersubjective reading is often available, in the sense that all participants in the conversation have access to the information source and are co-responsible for the inference made by the speaker (5):

(5) **Parece que** el viaje me hizo mucho bien; **según ellos** tenía mejor semblante

It seems that the journey did good to me; **according to them**, I had a better face.'
(Cornillie 2007 : 42)

13 Following this approach, he argues that a difference can be found between the two constructions by taking into account the dynamics of speaker-listener interaction. While this analysis could explain the difference between *parece que* and *parecer*+infinitive, it bears one issue: just like *parecer*+infinitive, in his analysis *me parece que* too expresses a subjective judgement (2007: 48); nevertheless, *me parece que* is more frequent than the other two constructions in the spoken corpus. Hence, this proposal based on the concept of (inter)subjectivity seems to make an incorrect prediction.

14 Summing up, these explanations highlight relevant differences between written and spoken situations (i.e., some peculiarities of speech, such as planning in on-line production and the dynamics of speaker-listener interaction). Nonetheless, they show some weaknesses. As a matter of fact, both the notion of writing and speech can be

problematic since they generalize over a number of very different communicative situations (e.g., both face-to-face dialogues and broadcasts are spoken situations). Due to their multifaceted nature, “writing vs. speech” does not provide us with a precise criterion to interpret the different distributions, nor a solid grid to decode the motivations underlying the occurrence of linguistic phenomena. Instead, attributing general properties to written and spoken language runs the risk of: 1) constituting a partial misunderstanding of the nature of our data (e.g., spoken data could be mainly composed of lowly dialogic situations); 2) being fallacious by overgeneralizing properties to communicative situations not characterized by them. We need a more systematic and fine-grained analysis of the features of spoken and written language to fully understand the motivations governing the distribution of these constructions.

3. Constructions in written and spoken language: making the case for text genres

15 As we said, the problems we encountered in the previous section can be attributed to an overly simplistic view of written and spoken grammar, which is not infrequent in studies on syntactic constructions.

16 Firstly, it must be noted that speech is often relegated in syntactic theories to an altogether marginal position, though there is relevant research investigating syntax in spoken language (e.g., Linell 1982; Blanche-Benveniste 1997; Voghera 1992; 2017; Lancheret-Dujour et al. 2019). This marginal position of speech in studies on syntax is accompanied by a tendency to take a partial or misleading view of the nature of utterances in the spoken language. For instance, when comparing the syntax of the spoken language to its written counterpart, it can be dangerous to rely on notions such as simplification or complexity (Ochs 1979; Berruto 1985). It should be acknowledged that spoken syntax is not “simpler”, but instead it is motivated by specific discourse conditions (Voghera 2017). Another misleading overlap that can be found in literature traces back the differences between written and spoken language to style variation (cf. the discussions in Koch & Oesterreicher 2001 and in Pistolesi 2015). This could suggest an improper association between (informal and formal) style and (written and spoken) mode. Instead, the whole range of styles can potentially be used in speech, although it is true that some manifest themselves more often in spoken situations (Cerruti 2009). Thus, the possible overlap between, e.g., informal style and spoken communication can only be verified through an analysis of the individual texts.

17 While these misconceptions are understandable in formal approaches to syntax that tend to marginalize linguistic performance, the risk of a “written language bias” (Linell 1982) has also been denounced in CxG (cf. Hilpert 2019). Nonetheless, many constructionist studies have begun to deal with constructions from (mostly dialogic) speech (e.g., Fried & Östman 2005, Linell 2009, Masini et al. 2018; Pietrandrea 2018; Geka et al. 2020), demonstrating the advantages of analyzing spoken language using constructions as units, rather than inherently written-centric notions such as ‘sentence’ (Auer & Pfänder 2011: 8). Much of this work on spoken language has dealt with phenomena typical of spontaneous dialogic speech, also due to their great distance from written grammar examples (e.g., co-construction phenomena, parentheticals, etc.). However, research on spoken communication should not be confined to the study of dialogic speech. This should be kept in mind when we contrast the frequency of choice of a construction in written and spoken corpora, because it could be problematic to resort to a ‘written vs spoken’ opposition in our interpretation of the data.

18 Looking for instance at the Italian tradition, an approach has often been to contrast the two poles representing prototypical written and spoken language, also called *scritto-scritto* ‘written-written’ (Lavinio 1990) and *parlato-parlato* ‘spoken-spoken’ (Nencioni

1976). These two poles are generally identified with the discursive situations that embody the greatest distance between spoken and written, such as spontaneous free conversations and academic prose. This approach, however, bears some problems. Firstly, it ignores text genres that are not easily framed by this opposition (Oesterreicher 1997; Calaresu & Palermo 2021). Secondly, it does not allow for adequate explanations: speech and writing are too composite (and often undefined notions) to be able to provide unitary accounts of phenomena. For instance, sometimes features typical of a prototypical pole are also generalized to text genres that do not (entirely) share them. Going back to the analyses on *parecer*-constructions in Section 2, they resort to explanations that are mainly associated with spontaneous dialogical speech, but that are not enough to be applied to spoken interactions as a whole. For instance, on-line planning can be found in different gradations in different types of speech, and, at the same time, there are highly planned types of speech (i.e., spoken monologues). To avoid such problems, we need to profile the text genres that make up our corpus.

19 It could be argued that a genre-based categorization of texts is not sufficiently precise to determine, e.g., their level of formality or their function. As mentioned for speech and writing, there is no one-to-one relationship between textual genres and degree of formality, and, in addition, the same textual genre is not homogeneous (Shahrokny-Prehn & Höche 2011). Furthermore, the same text can include sections pertaining to different types (e.g., descriptive, narrative, scenic, etc.) (Lavinio 2000). However, we claim that a categorization in text genres based on external criteria (their material appearance as fiction books, or as press articles, etc.) has the advantage of being carried out quite objectively and allows us to detect interesting tendencies. Thus, we can see this approach as a middle ground between precision and practicality in the analysis.

3.1 Enriching CxG to address non-genre specific constructions: continuum-based approaches

20 The relevance of the relationship between constructions and text genres is not new to constructionist approaches. Many studies have highlighted the necessity of including text genres within the linguistic knowledge of speakers - and thus, within the representation of constructions (e.g., Bergs 2010) - by emphasising how context is able to determine the acceptability of otherwise agrammatical constructions (e.g., Östman 2005; Matsumoto 2010; Nikiforidou 2010, 2021; Ruppenhofer & Michaelis 2010). This has led to considering text genres as frames that license certain constructions (Östman 2005) and interact with processes of linguistic change (e.g., Fried 2009), or even to modeling text genres as constructions (Hoffmann & Bergs 2018).

21 While many of the studies in CxG have dealt with genre-based constructions, our focus is slightly different but complementary. Indeed, the constructions with *sembrare* are part of what Hilpert calls “normal syntax” (2019: 68), and in principle, they are less idiosyncratic than structures acceptable only in specific text genres. However, as suggested by Fillmore et al. (1998), one of the goals of CxG is also to account for more familiar structures. By making the case for the relevance of a contextual analysis of “normal” constructions, we aim to strengthen the idea that speakers' knowledge about a construction includes information pertaining to the external context.

22 As we have mentioned, our aim is to consider text genres according to their specific features. However, we also want to identify general properties that allow us to provide explanations going beyond individual text genres since our constructions tend to be grammatically acceptable in (almost) every environment.

23 It may therefore be useful to refer to scalar, or continuous, models of written and spoken texts (e.g., Biber 1988; Koch & Oesterreicher 2001). These models identify text genres by arranging them on a gradient of one or more properties (e.g., involvement, planning, communicative distance, etc.), chosen on theoretical or empirical grounds. In particular, more theoretically oriented approaches define the gradation of these

properties based on certain structural parameters of individual textual genres. For example, in the classification of communicative situations in the LIP spoken Italian corpus (De Mauro et al. 1993) a number of criteria are used, such as: presence or absence of interlocutors; direction of communication; synchrony between production and reception; constraints on turn-taking (Voghera et al. 2014). A situation that displays presence of interlocutors and bidirectional, unconstrained flow of information can be defined as highly dialogic. As for the written mode, there is definitely more homogeneity of communicative situations: interlocutors are not present, and reception and production are asynchronous (Voghera 2022). However, a gradation is provided by other features of text genres: for example, fiction prose contains dialogic sections, and, to a lesser extent, press does too, since it contains interviews; for this reason, some written genres are closer to some spoken genres in terms of the linguistic phenomena they present (Voghera 2017).

24 It is worth clarifying that the use of these parameters is not exhaustive of the speaker's knowledge of a text genre, and only describes a limited part of it (cf. the discussion in Nikiforidou 2016). However, such classifications of texts prove to be useful in detecting trends in the distribution of constructions: for instance, even data from studies that do not focus on contextual analysis show that a number of phenomena behave accordingly to these scales, such as the frequency of topicalizations (Frascarelli 2003) and of verb-particle constructions (Iacobini 2008) in spoken Italian, or the collocational profiles of modal auxiliaries in English (Hilpert 2016).

25 In particular, Voghera hypothesizes a scale of written and spoken Italian texts from the observation of the frequency of use of nouns and verbs and based on theoretical considerations (2017: 143-147). The text genres are ordered according to two criteria: the degree of dialogicity and the degree of planning (Figure 1).

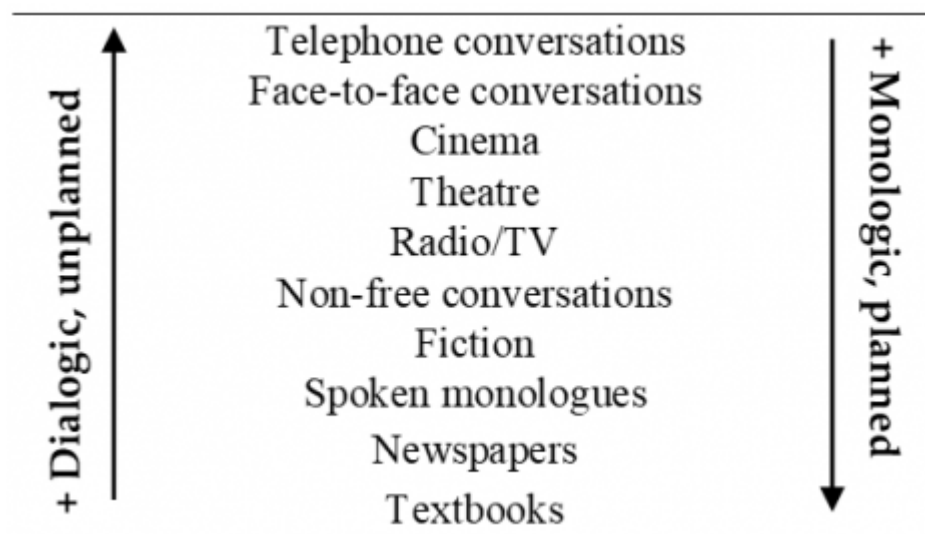


Figure 1. The dialogicity and planning scale of text genres described in Voghera (2017)

26 The use of two criteria is crucial, because they jointly contribute to the positioning of texts on the scale. For instance, planning possibility is generally the distinguishing factor between spoken genres (except for Spoken monologues, which are planned but also monologic), whereas between the written genres (which are generally equally planned) the degree of monolocicity is the most relevant aspect (e.g., Fiction is closer to the spoken genres because it contains dialogic sections) (2017: 146).

27 Interestingly, this scale corresponds (with minor differences) to the first dimension of variation of English observed by Biber (1988) (and see also Oesterreicher 1997: 195). Dimension 1 is considered among the most important of the six found by Biber (Raible 2019) and is one of the two Biber (2014) found to account for linguistic variation cross-linguistically. The different interpretations of this dimension led Biber to emphasize different properties of the two poles of the continuum, defined as:

- interactive vs edited texts (Biber 1986b, cf. also real-time vs planned in Biber 2014);
- involved vs informational productions (Biber 1988);
- oral vs literate discourse (Biber & Conrad 2009).

28 Despite the differences, most of these characteristics fall within the two criteria defined by Voghera, which in turn depend on the two contextual coordinates that distinguish written and spoken situations, namely temporality (on which the possibility of language planning depends) and the presence/absence of interlocutors (dialogicity/monolocicity). We will use this approach precisely because it allows us to look at the text genres separately but ordered according to fundamental features of communicative situations, while also preserving a general distinction between spoken and written modes, which, as we already saw (Section 2), is relevant for the distribution of our constructions.

4. Analysis: linguistic and extra-linguistic factors in the choice of *sembrare*-constructions

29 Our analysis focuses on three constructions with *sembrare* to delimitate the investigation to a common formal and functional space, namely *sembrare*-constructions that govern a complement clause, and that in principle should be interchangeable from a truth-conditional point of view. Our aim is to assess the influence of text genres - and thus of communicative situations - on the use of *sembrare*-constructions. We will address how relevant text genre is as a constraint in predicting the speakers' choice by comparing it with other previously studied constraints on *sembrare*-constructions. In fact, as we mentioned in the Introduction, some linguistic variables have been proven to affect the choice between *sembrare*-constructions. However, by looking at the distribution between written and spoken corpora in Italian (Section 2), there seems to be a quite strong opposition between the context of production of the constructions under study. In light of these observations and of the discussion in Sections 2-3, we hypothesize that:

- extra-linguistic factors contribute to determining the choice between *sembrare*-constructions;
- a continuum-based model of text genres explains variability better than simply contrasting written and spoken language.

4.1 Dataset

30 The dataset is the one used in Pisciotto (2023), and it is annotated for a number of linguistic variables (see Table 4 in Section 4.2) previously analyzed by means of monofactorial tests (see Section 1). For our purposes, a new variable was added, that is the text genre in which the occurrence is found. The dataset contains 405 total occurrences of the three constructions *mi sembra che* (6), *sembra che* (7) and *sembrare*+infinitive (8), coming both from written and spoken corpora. We didn't take into account *sembrare*+infinitive with the overt experiencer, because of the lack of occurrences of this construction (see footnote 4).

(6) **Sembra** **che** Carla abbia ragione.

seem.IND.PRES.3SG COMP Carla have.SBJV.PRES.3SG reason.SG

'it seems that Carla is right.' (CORIS MONITOR2005_07, NARRAT)

(7) **mi** **sembra** **che** non sia male
to-me seem.IND.PRES.3SG COMP NEG be. SBJV.PRES.3SG bad.SG
I' idea [...]
ART.DEF.SG idea.SG
‘it **seems to me that** it is not a bad idea’ (LIP, MB11)

(8) **la** politica [...] **sembra** **aver** **smarrito** ogni
ART.DEF.SG politics.SG seem.IND.PRES.3SG have.INF lose.PTCP any
senso etico
sense.SG ethical.SG
‘politics **seems to have lost** all of its ethical sense’ (LIP, RD14)

31 The written data were extracted from CORIS (Rossini Favretti et al. 2002). CORIS is a balanced and representative corpus of Standard written Italian, regularly updated through the addition of monitor corpora. At the time when the research was carried out, it contained 150 Mw (monitor corpus 2017). We extracted a random sample of 622 occurrences of the verb *sembrare*. Among these occurrences, 216 were coded as *sembrare*+infinitive, *mi sembra che* or *sembra che*.

32 As for spoken Italian, two corpora were used to gather an amount of data similar to the written ones: LIP (De Mauro et al. 1993) and KIParla (Mauri et al. 2019). At the time when the research was carried out, the two corpora put together contained 1.5 Mw. The total occurrences of *sembrare* were 950, but only 189 of them were coded as *sembrare*+infinitive, *mi sembra che* or *sembra che*.

33 As we mentioned, the occurrences had previously been annotated. The linguistic variables included are related to the (“raised” or “unraised”) subject slot filler, and were chosen following previous research on other languages (Davies 1997; Achard 2000; Grimm 2010). We briefly summarize in Table 2 the coded variables in the dataset.

Variable	Levels
Length (of the subject)	n. of words (0,1,2,3,4, etc.)
Animacy (of the subject)	Human, Animate, Non-Animate
Referent accessibility (of the subject)	Old, Mediated, New
Type of anaphoric link (between the subject and its antecedent, cf. Nissim et al. 2004)	Old_identity, Old_relative, Old_generic, Old_ident_generic, Old_general, Old_event, Mediated_bound, Mediated_general, Mediated_event, Mediated_aggregated, Mediated_func_value, Mediated_set, Mediated_poss, Mediated_part, Mediated_sit, New

Table 2. Coded variables in the dataset

34 Two of these linguistic variables proved to be significant constraints on the constructions with *sembrare*, namely the length of the subject and the type of anaphoric link⁶ (see Section 1).

35 Subsequently, one new variable was added: we annotated the occurrences in our dataset for their text genre of origin. For the written occurrences, we used the metadata available in the corpus. CORIS includes six text genres, defined mainly according to external criteria: Fiction, Newspapers, Magazines, Academic prose, Legal prose, and Ephemera (handwritten, printed and electronic texts characterized by their short

permanence, e.g., instructions, private letters, blogs, etc.). However, for the purposes of our analysis, we excluded the six occurrences coming from Ephemera, given the strong heterogeneity of the text genres and purposes (and consequently the difficulty in cataloguing them). We also combined the occurrences from Magazines with those from Newspapers, considering them as part of the macro-category Press.

36 As for the spoken corpora, there is a discrepancy in the metadata of the two corpora (LIP and KIParla). The communicative situations included in KIParla are: Free conversations, Group conversations (with a moderator), Semi-structured interviews, Student receptions, Exams, and University lessons. Instead, we have a wider range of communicative situations in LIP, categorized in five macro-groups of communicative situations: Free face-to-face conversations, Telephone conversations, Non-free face-to-face conversations, Monologic speech, and Radio/TV speech. Thus, we chose LIP subdivision to code our occurrences. To do so, we tagged KIParla communicative situations following the criteria for the subdivision of LIP texts (De Mauro et al.1993). We considered: Free conversations and Student reception (KIParla) as Free face-to-face conversations (LIP); Group conversations, Semi-structured interviews and Exams (KIParla) as Non-free face-to-face conversations (LIP); University lessons (KIParla) as Monologic speech (LIP).

37 The resulting classification is composed of nine text genres, as illustrated in Table 3 along with the distribution of the constructions.

	<i>sembra che</i>	<i>mi sembra che</i>	<i>sembrare</i> +infinitive	Total
Academic Prose	11	1	31	43
Fiction	31	12	33	76
Free conversations	13	33	3	49
Legal Prose	1	0	9	10
Monologic speech	15	6	13	34
Non-free conversations	29	40	7	76
Press	16	10	53	79
Radio/TV speech	4	9	2	15
Telephone conversations	2	12	0	14
Total	122	123	151	396

Table 3. Distribution of *sembrare*-constructions in the coded text genres

38 In order to test hypothesis (b) (see Section 4), we treat text genres as an ordinal independent variable, as ranked in Voghera's (2017) scale (Section 3.1). We coded them as ranks going from 1 to 9, the 1st representing the most dialogic and less planned genre (Telephone conversations), and the 9th representing the less dialogic and most planned genre (Legal Prose). As can be seen in Table 3, some genres are absent in our corpora (Cinema and Theatre, cf. with Figure 3 in Section 3.1). Furthermore, we added Legal Prose, absent in Voghera's classification, as the highest-ranked genre, accordingly to the position of "official documents" in Biber's (1988) Dimension 1.

4.2 Statistical methods: Conditional Inference Trees

39 Our analysis investigates the relevance of contextual constraints, and in particular text genre, in the choice of constructions with *sembrare*. We will try to predict speakers' choice by also considering other linguistic variables, in order to understand the weight

of extra-linguistic variables as compared with linguistic ones in predicting the choice of one of the three constructions. To do so, we fit a Conditional Inference Tree (CIT) (Hothorn et al. 2006). CITs are non-parametric and recursive partitioning methods that are helpful to model the relationship between a dependent variable and various predictors. They do so by splitting the data according to the predictor that helps best to distinguish between the values of the dependent variable, and by repeating this process on the subsets until no further split can be made (for a more comprehensive introduction, cf. Levshina 2021). CITs have been fruitfully used to study and model linguistic variability, and also in studies on alternating syntactic constructions (e.g., Szmrecsanyi et al. 2016).

40 We fitted the model using the *partykit* package for R (Hothorn & Zeileis 2015), including the corpus (Written/Spoken) and text genre of origin, and the linguistic variables: the length (*S_Length*), the type of anaphoric link (*TypeOfAnaphor*), the animacy (*Animacy*) and the referent accessibility of the subject (*Accessibility*). The dependent and the independent variables were converted to factor, except for *S_Length* and *TextGenre_Ordinal* (i.e., text genres expressed as a 1-9 scale), which were treated as a numeric. Given the relatively small size of our sample we employed Monte-Carlo method as test type, and the minimum sizes for the split and for the buckets were reduced to 5⁷. Subsequently, we also fitted two other trees for a deeper understanding of our results: in the first, *TextGenre_Ordinal* was removed (i.e., employing as contextual information only *Corpus*); in the second, no contextual information was included, and only linguistic variables were analyzed. In the Results Section we will mainly focus on the tree with all the predictors included and briefly show some differences with the other two models; however, the full output of the three models can be found in the online repository (see the Appendix).

5. Results

5.1 Descriptive statistics: text genres as a scale of dialogicity and planning

41 The descriptive analysis focuses on whether there is a relationship between the choice of one of the three constructions and the dialogicity and planning of the texts. In Table 4 we show the percentages referring to the proportion of each construction in each text genre (as can be seen, the values of the three constructions add up to 100 for each text genre). Text genres were ordered according to Voghera's (2017) scale.

	<i>sembra che</i>	<i>mi sembra che</i>	<i>sembrare</i> +infinitive	Total
1 (Telephone conversations)	14.29%	85.71%	0%	100%
2 (Free conversations)	26.53%	67.35%	6.12%	100%
3 (Radio/TV speech)	26.67%	60%	13.33%	100%
4 (Non-free conversations)	38.16%	52.63%	9.21%	100%
5 (Fiction)	40.79%	15.79%	43.42%	100%
6 (Monologic speech)	44.12%	17.65%	38.23%	100%
7 (Press)	20.25%	12.66%	67.09%	100%
8 (Academic Prose)	25.58%	2.33%	72.09%	100%
9 (Legal Prose)	10%	0%	90%	100%

Table 4. Proportions of the constructions for each text genre, ordered according to Voghera's (2017) scale.

42 As can also be seen in Figure 2, the choice of *sembrare+infinitive* tends to increase almost monotonically with the higher monologicity and planning of a text, while the percentages show the opposite situation with *mi sembra che*. In contrast, *sembra che* does not show a clear trend on this scale. Furthermore, it can be noted that monologic speech shows percentages closer to written genres than to spoken ones.

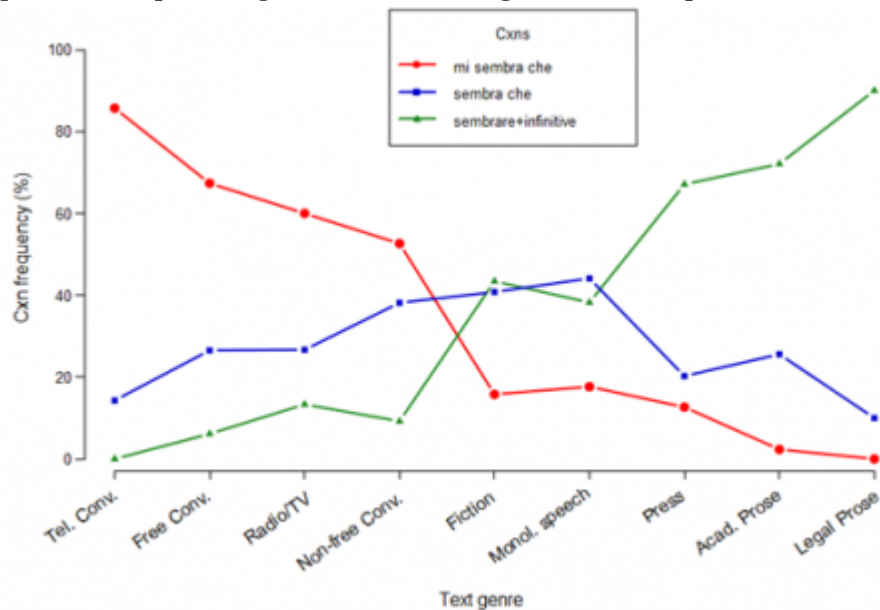


Figure 2. Proportions of the constructions for each text genre, plotted against Voghera's (2017) scale.

43 In order to test the association between the increase of dialogicity and planning and the use of our constructions, we calculate Kendall's tau correlation coefficient (τ) in R, repeating the test for each of the three constructions. The dependent variables of the tests are the proportions (in percentage) of the construction within each text genre in our dataset, while in all the tests, the ordinal independent variable is represented by Voghera's (2017) scale, expressed as ranks going from 1 to 9. The results of the tests confirm our observations:

- the frequency of *sembrare+infinitive* has a very high positive correlation with the scale ($\tau = 0.96$, p-value < 0.001);
- the frequency of *mi sembra che* has a very high negative correlation with the scale ($\tau = -0.96$, p-value < 0.001);
- the test on *sembra che* shows no significant correlation between the construction and the scale ($\tau = -0.10$, p-value = 0.7).

44 Thus, it could be argued that, at least for *mi sembra che* and *sembrare+infinitive*, a scalar model of text genres could constitute a useful point of view to describe variation.

5.2 A tree model of *sembrare*-constructions alternation

45 The accuracy of the model, though not very high (0.61), proved to be significantly higher than the baseline for the model (No Information Rate = 0.38, p-value = 2.2e-16).

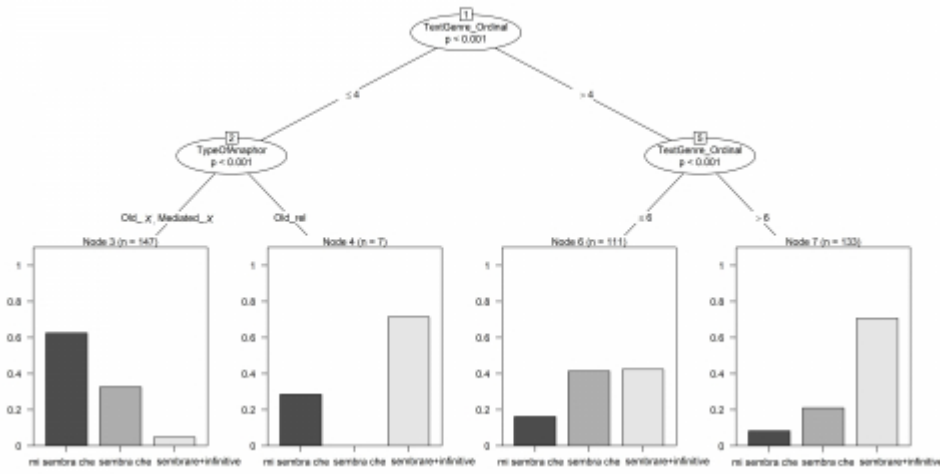


Figure 3. Conditional Inference Tree distinguishing between the three sembrare-constructions

46 As we see in Figure 3, the first split in the CIT is based on dialogicity and planning (Node 1), reflecting the distributions seen in Section 5.1. On the left, we find the texts ranked up to 4 (the most dialogic and unplanned genres), generally predicting *mi sembra che*. The only exception in this subgroup is a further split is made based on the Type of Anaphor: if the subject is a relative pronoun (Old_rel) the model predicts the choice of *sembra+infinitive* (Node 4). On the right, we find instead a further split based on text genres: in Node 7 the ranks higher than 6 (the most monologic and planned genres) predict *sembra+infinitive*, while in Node 6 the ranks 5 and 6 are grouped together, showing a subgroup with almost a 50/50 distribution between *sembra che* and *sembra+infinitive*. Interestingly, written and spoken occurrences are generally distinct (thus confirming the tendencies observed in Section 2). This is not strange, since the scale model we employed keeps an overall distinction between oral and literate discourse (Section 3.1). However, notable exceptions are monologic speech, which is grouped together with written genres (as predicted by our scale), but also fiction, which is later separated from the rest of written genres, and grouped together with monologic speech, since the two yield a similar distribution. We then checked the conditional variable importance, to assess the relevance of the two variables selected for the splits (Figure 4).

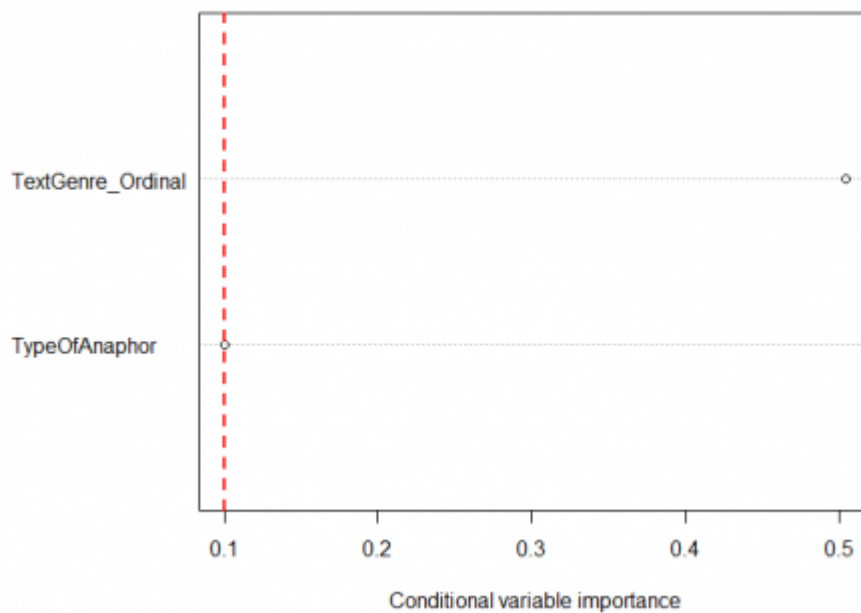


Figure 4. Relative importance of the variables for distinguishing between the three sembrare-constructions

As shown by Figure 4, in this model the scale of dialogicity and planning proves to be the most important variable in predicting the choice between *sembra che*, *mi sembra che* and *sembrare*+infinitive.

In order to better understand why Corpus was excluded from the tree, we fitted a model without TextGenre_Ordinal. Interestingly, such model yields a similar accuracy; however, the analysis of the relative importance of the variables shows that Corpus has basically the same importance as the Type of Anaphor (both 0.44) (note that in the previous model the difference is much higher: TextGenre_Ordinal = 0.50, TypeOfAnaphor = 0.09) (Figure 5). This means that, as a variable, when we only employ a “written vs spoken” distinction, we need to appeal more to linguistic predictors, since Corpus as a variable has less predictive power than TextGenres_Ordinal, and that extra-linguistic context has a higher predictive value when expressed as a scale of text genres.

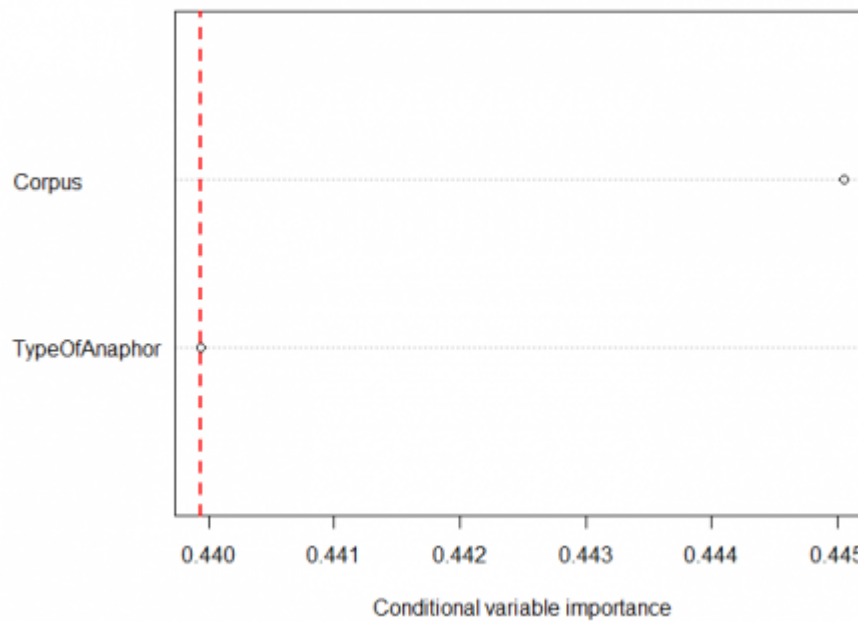


Figure 5. Relative importance of the variables for distinguishing between the three sembrare-constructions (TextGenre_Ordinal removed).

Finally, we contrasted such models with a model without contextual informations. While this last model, which relies on Type of Anaphor and Subject Length, has predictably a lower accuracy (0.48) than the previous two, it has interestingly a higher sensitivity in predicting *sembra che*. In fact, the relatively low accuracy of the “context-based” models is due to the difficulty in predicting *sembra che*. If we look at the model with all the variables, the sensitivity of for this class is very low (0.07), in contrast to *mi sembra che* (0.74) and *sembrare*+infinitive (0.92), and the values are similar in the “corpus-only” model. Instead, the “linguistic-only” model has a somewhat higher sensitivity for *sembra che* (0.18) and a lower sensitivity for *mi sembra che* than the other models (0.31), pointing at possible drawbacks in considering only extra-linguistic cues for some constructions.

6. Qualitative analysis: the influence of dialogicity and planning on *sembrare*-constructions

The results showed that *mi sembra che* and *sembrare*+infinitive are mainly predicted by contextual variables, while linguistic ones play a secondary role. Furthermore, the

contextual variable selected by the model turned out to be the scalar ranking of text genres. In the following section, we will suggest an interpretation of the results of our tests through a qualitative analysis of the occurrences in the dataset. This analysis will be based on the two criteria of Voghera's (2017) scale: dialogicity (Section 6.1) and planning (Section 6.2).

6.1 Variability in *sembrare*-constructions: the role of speaker's involvement

51 As we saw in Section 5.2, the two constructions whose use is most context-dependent are *mi sembra che* and *sembrare*+infinitive. How can we account for their variation across text genres? The first factor to consider is dialogicity, and in particular the relationship between speaker and listener.

52 Firstly, we can note that both of these constructions generally express the speaker's point of view on the likelihood of some state of affairs. This is quite evident with *mi sembra che*, in which we find a 1st person experiencer (that is, the speaker) in 84% of the occurrences. As for *sembrare*+infinitive, we can apply a test (cf. Cornillie 2007; Musi 2015) which shows that the use of this construction would not generally be felicitous if the speaker did not recognize the opinion expressed as their own (9)⁸.

Per lo meno a partire da "I clowns" e poi man mano con un'accentuazione crescente fino a
(9) "La voce della luna", **Fellini sembra essere approdato** ad una grande forma di saggezza [...].

? ...però non credo sia così.

'At least starting from 'The Clowns' and then gradually, with increasing emphasis until 'The Voice of the Moon', **Fellini seems to have approached** a great form of wisdom [...].'
(CORIS 1980_2000, PRACC)

'? ...but I don't see it that way'

53 Although both *sembrare*+infinitive and *mi sembra che* express a subjective judgement, their formal difference unveils a difference in the construal of the situation. Indeed, in *mi sembra che* the speaker is explicitly marked as the conceptualizer of the judgement, whereas in *sembrare*+infinitive the speaker is 'hidden'.

54 We can link this difference to a different degree of speaker involvement, which is a typical feature of spoken language, but mostly of dialogic situations (Chafe 1985). Involvement, according to Chafe, manifests itself as the speaker's involvement towards themselves (ego involvement)⁹ but also towards their interlocutors (interpersonal involvement). Ego and interpersonal involvement favour an explicit marking of the speaker and of the interlocutors by means of personal pronouns, as in the case of *mi sembra che*. Moreover, ego involvement determines a more explicit marking of the speaker's mental processes. In our case, this occurs when it is overtly expressed that it is the speaker the conceptualizer of a judgement over a state of affairs (1985: 116-117). For instance, in (10) the speaker is clearly expressing that the statement is based on their personal opinion:

mi sembra che [...] l'idea della tolleranza della cultura l'idea del rispetto eh sia comunque
(10) un elemento che la [=il multiculturalismo] qualifica [...] come un passo avanti rispetto a ai precedenti orientamenti

'it seems to me that the idea of tolerance of the cultures, the idea of respect are nonetheless elements that qualify it [=multiculturalism] as a step forward as compared to previous approaches.' (LIP, RC2)

55 On the contrary, *sembrare*+infinitive not only disguises the fact that the speaker (/writer) is the conceptualizer, but also presents a focus on a particular element (the subject) of the proposition on which the judgement is expressed. If we look at the

example (11) the writer seems to convey that “it is the oscillation of the viral load that gives the impression to have an influence, not me to suggest it”:

- (11) l'oscillazione della carica virale sembra influenzare maggiormente l'ondulazione della conta linfocitaria [...].

‘the oscillation of the viral load seems to have more influence on the undulation of the lymphocyte count [...]’ (CORIS MONITOR2001_04, PRACC)

56 In other words, prominence is given to the content of the proposition - or at least to a part of it - and not to the cognitive process behind the judgement (cf. the analysis in Langacker 1995). Thus, it is not the subjective vs. intersubjective difference that guides the speaker's choice between constructions with *sembrare* (cf. Section 2), but rather the choice to present a judgement about some state of affairs in an involved or informational manner.

57 Hence, *sembrare*+infinitive is favoured in genres in which, as the speaker and the interlocutors are not present, there is “little overt acknowledgement of the personal feelings of the speaker/writer” (Biber 2014). Moreover, within some of these “informational” genres (Academic Prose and Legal Prose) it would be awkward (though still grammatical) to use *mi sembra che*, unless licensed by a particular context.

58 Finally, what place can we assign to *sembra che* in this analysis? This construction appears more in the “in-between” genres, which are not informational enough to foreground some topical information, but neither are dialogical (or at least not in the strict sense of the term, see the discussion in Section 7.1). Thus, its increase in frequency compared to the other two constructions is not due to inherent properties. Instead, it seems to be an “unmarked” variant with respect to speaker’s involvement, since neither the experiencer nor propositional information are foregrounded, and its choice could be favoured by non-contextual factors, such as linguistic constraints.

6.2 On-line and offline production constraints, between form and function

59 Planning possibilities in the different text genres too may influence the variability in the choice between the constructions. It has been claimed that *that*-complements “mark informational elaboration that is produced under strict real-time constraints” and thus permit the on-line elaboration of information that cannot be adequately planned (Biber 1988: 113, 195). Thus, in our case, the choice of *(mi) sembra che* over *sembrare*+infinitive could be due to planning constraints: in fact, in on-line production the speaker would need to already have the subordinate proposition, or at least its subject, available in order to be able to produce the element in the first position in *sembrare*+infinitive. And indeed, in about half of the cases in which *sembrare*+infinitive is used in the spoken data, the subject is favoured to be in the first position because it is a relative pronoun (12), and thus the speaker finds the subject of *sembrare*+infinitive already structurally available.

- (12) [...] c’ è la musica **che sembra**
there be.IND.PRES.3SG ART.DEF.SG music.SG REL seem.IND.PRES.3SG
avere solo un valore estetico [...]
have.INF only ART.INDF.SG value.SG aesthetic.SG

‘there is music, **that seems to have** only an aesthetic value’ (KIParla, TOD1002)

60 On the contrary, it is known that infinitive subordinates are generally more characteristic of written genres (Voghera 1992: 237-239; Kirk 1997). As noted by Davies (1997a; 1997b, cf. Section 2), in written registers it is preferred to use more integrated structures, such as *sembrare*+infinitive, which tends to be more “monoclausal” than its

counterpart with finite verb complementation. We could link this tendency to the search for a higher lexical density (i.e., the ratio between content words on the total number of words) in more informational texts (Voghera 2017). This can be seen in (13a-b), where we repeat the examples (1a-b) (content words are in bold):

- (13) a. **Sembra** che **Gianni dorma**. (*sembra che*, lexical density : $3/4 = 0.75$)

'It seems that Gianni is sleeping.'

- b. **Gianni sembra dormire**. (*sembrare+infinitive*, lexical density: $3/3 = 1$)

'Gianni seems to be sleeping.'

61 The need for lexical density can be also motivated by specific external conditions of text genres: for instance, in newspaper articles, it is generally preferred to omit functional words to achieve more brevity and conciseness (e.g., cf. Szmrecsanyi & Hinrichs 2008). In more planned genres, it would be possible to achieve such brevity through the possibility to edit linguistic productions.

62 Furthermore, the possibility of planning a text not only intersects with formal aspects of the constructions, but also with functional (evidential and epistemic) ones. For instance, in conversational situations speakers present fewer sources of evidence for their statements than in written and planned registers (that is, their use of epistemic markers often does not point to a specific source of information, see Chafe 1986 and Pietrandrea 2018). In most of our occurrences, *mi sembra che* expresses judgements that seem to be based entirely on the speaker's belief (14) (as we saw, this is formally marked by the foregrounding of the experiencer) or on sources such as memory (15), which are generally less reliable but more accessible in unplanned interactions (since memory is a speaker-internal source):

- (14) *mi sembra che non sia male l'idea di fare un servizio solo che bisognerebbe farlo in modo un po' articolato capisci*

'it seems to me that it is not a bad idea to do a report, but it should be done in a somewhat articulate way, you see' (LIP, MB11)

- (15) TOR001: *tu c'eri all'addio al calcio di maldini*

'where you there when Maldini played his last football game ?'

TOI003: *no non c'ero // [...] mi sembra che era **mi sembra che** era alle tre di pomeriggio quindi o giocavo*

'no, I wasn't there [...] it seems to me **it seems to me that** it was at three in the afternoon so either I was playing or..' (KIParla, PTA003)

63 It could be that, because of its epistemic meaning, *mi sembra che* is preferred in less planned interactions and dispreferred in more edited texts, in which the speaker is more careful and certain about their sources. However, it is also true that this construction can also appear with indications of the speaker's sources of evidence within more planned genres (e.g., Fiction) (16):

- (16) *Mi sembrava ci fosse fuliggine ovunque, e da ogni angolo un puzzo diffuso di carbone.*

'It seemed to me there was soot everywhere, and a widespread stench of coal coming from every corner.' (CORIS MONITOR2014_16, NARRAT)

64 In (16) the writer focuses on a perceptual datum and draws an inference on it to make up their belief about the state of affairs. While in the literature it is claimed that inferences on perceptual data are expressed by *sembrare+infinitive* (cf. Salvi 1991), in this case *mi sembra che* can be used to refer to the same kind of knowledge acquisition.

65 This could hint at a degree of functional flexibility of *mi sembra che*: its functional information can be adapted to the specific conditions of different text genres and thus become more similar to the one expressed by the two other competing constructions. Nonetheless, it is difficult to imagine *mi sembra che* used in highly monologic genres,

and overall, the higher frequency of *mi sembra che* in unplanned genres shows that the most “natural” niche for this construction are dialogic and unplanned texts.

7. General discussion

66 In this section we will discuss the implication of our results on the status of text genres in the study of alternations. First, we will address to what extent text genres can explain variation (Section 7.1) and then, we will briefly sketch a possible way to include contextual information in the representation of constructions (Section 7.2).

7.1. On text genres: what they can and cannot explain

67 The model we fitted shows that contextual variables have the highest weight in predicting the choice between *sembra che*, *sembrare*+infinitive, and *mi sembra che*. Furthermore, both statistical results and the qualitative analysis showed the advantages of assuming a scalar perspective to text genres. These facts allow us to draw some considerations.

68 Above all, we should stress the relevance of investigating variation across text genres and not only by looking at the differences in written and spoken language. In fact, even though the CIF yields three different groups, it could be argued that the main split in the model reflects somehow a difference between a “spoken-pole” (dialogic speech, including broadcast) and a “written-pole” (basically non-fiction prose), thus preserving the intuition that the choice is a matter of speech and writing. However, our analysis provided at least three different pieces of evidence in favour of a text genre-based analysis, the first two concerning the explanatory power of text genres both in our predictive (a) and in our descriptive (b) analyses, and the third regarding the peculiar position of fiction and monologic speech, that do not behave accordingly to their corpora of origin (c). We briefly focus on them one by one:

- a. *Variable importance in the model*: text genre is the contextual variable selected by the model instead of the corpus of origin (despite the dimension of our sample makes it difficult to catch all the possible differences between nine different categories). As shown by the relative importance of the variables, this is due to the fact that our scale of text genres explains more variation taken alone than written/spoken distinction, that is: if we assume that speakers have knowledge of the features of communicative situations, less strictly linguistic information is needed to explain the choice of the constructions.
- b. *Explanatory power of the scalar approach*: a finer-grained observation of the quasi-monotonic trend for *sembrare*+infinitive and *mi sembra che*, as provided by data visualization, should be taken into account (Section 5.1). For instance, it is understandable why the CIT keeps together the ranks from 1 to 4, since they all show a high number of *mi sembra che* in proportion. However, even though the scale we chose preserves a distinction between written and spoken language (Biber & Conrad 2009), a “written vs spoken” opposition does not necessarily entail a monotonic distribution of the constructions. An example is in Pesciotta & Masini (accepted), regarding the distribution of analytic versus synthetic psych-predicates in Italian. In that case, while there is a split between written and spoken corpora, there is not a clear scale between text genres, with a more skewed behaviour (i.e., broadcast speech strongly differs from the distribution of spoken genres and behaves more similarly to non-fiction prose): in such a case, providing our scale as a variable would have resulted in a less clear picture. Moreover, as shown by the qualitative analysis, our scalar model provides us with solid interpretative criteria since they are built in the ranking of text genres.

Such criteria are not inherently provided by composite categories as written and spoken situations, while in our case can be safely assumed as explanatory.

- c. *The idiosyncratic behaviour of fiction and monologic speech*: finally, as we mentioned, there is an interesting split in our model, since fiction and monologic speech are separated from their respective corpora and are grouped together. Thus, we could argue that the CIT provides at least a three-way split, improving the binary alternative between written and spoken language: dialogic genres, monologues and fiction prose, and non-fiction prose. This should alone provide a reason to prefer a text genre representation and to avoid a “written-pole” perspective, since it is at least debatable that fiction prose is a less prototypical written genre than academic or legal prose. This grouping is consistent with the scale we adopted, since monologic speech is positioned between fiction on the one hand and other written genres on the other.

69 One possible question could be thus why fiction and monologic speech are lumped together, that is, whether this third pole is motivated by inherent similarities between the two genres. Rather than forming a functionally motivated class (e.g., informational vs. involved, narrative vs. nonfictional), we might hypothesize, based on our scale, that their closeness depends on their hybrid nature when it comes to planning and/or dialogicity. Starting with planning, obviously fiction (like the other written genres in our sample) is a highly planned and edited genre. Monological speech is less planned than fiction, yet its greater degree of planning than the rest of spoken genres favours the use of a set of patterns and strategies similar to those of written genres (for instance, the use of *sembrare*+infinitive can be seen as a topic-marking strategy in the absence of context, cf. Section 6.1 and Pesciotta 2023). The use of these cohesion strategies is also due to monologicity: despite the presence of an audience, textual cohesion is not constructed with interlocutors intersubjectively (and this explains the lower use of *mi sembra che*) (Thompson 1994). Turning then to dialogicity, we note that fiction tends to be monological, but has dialogic sections of varying width (Voghera 2017). In addition, there is a tendency in many fiction productions to imitate dialogic/colloquial speech (Calaresu 2005). This may increase the frequency of some structures which are typical of conversational genres. In turn, monologic speech sections in corpora contains dialogic sections too, e.g. classroom interaction during class. So, in both genres dialogicity and monologicity are present in a more hybrid way than in other genres. For these reasons, it is plausible that the two genres are distinguished from conversational speech and non-fiction prose.

70 That being said, it is important to highlight that our results do not imply that there are no other differences between the constructions apart from their context of use, nor is our aim to select the model including text genres only because it factors out the impact of linguistic variables. For example, the model itself highlights the relevance of a constraint found in Pesciotta (2023), namely that on the preference in the choice of *sembrare*+infinitive in subject relative clauses. Moreover, monofactorial tests allow us to find other possible linguistic niches for some of our constructions (Pesciotta 2023): an example is 1st and 2nd person subjects (cf. *Old_general* in Table 2), which are only used with *sembra che* and *mi sembra che*, while are generally not used with *sembrare*+infinitive (in our dataset *sembrare*+infinitive never appears with a 1st or 2nd person subject). However, said constraints can only account for a small part of the variation: the contexts of relative subject clauses and 1st and 2nd person subjects concern only 35 and 26 occurrences respectively. In particular, 1st and 2nd person subjects are grouped by the model with all the other types of anaphoric links in the “spoken” subset, and do not stand out since *mi sembra che* and *sembra che* already constitute the most frequent class in that subset. Hence, we could consider this constraint to be nearly categorical but not often available as a cue for predicting the choice of the construction¹⁰. In contrast, the production context allows to predict a fairly large part of variation and is therefore more important in the probabilistic choice between these variants. Nonetheless, this does not preclude the necessity to identify

other possible linguistic factors (e.g., features of the subordinate verb), also because the accuracy of the model is not very high.

71 Moreover, as we saw, extra-linguistic factors do not always contribute to explain variation: it is the case of *sembra che*, whose distribution is not adequately captured by our models. The behaviour of *sembra che* could be interpreted in at least two different ways: on one hand, we could assume that *sembra che* is some sort of “unmarked” construction, that is, a construction in which no element is foregrounded (see Section 6.1). However, such unmarkedness clearly does not entail a higher frequency with respect to the other two constructions, nor we could assume that *sembra che* is somewhat more “basic” or canonical than *sembrare*+infinitive and *mi sembra che*. On the other hand, as suggested by one of the anonymous reviewers, we could thus think of *sembrare*-alternation as some sort of “nested” alternation: while *mi sembra che* can be clearly distinguished from the other constructions solely on contextual grounds (and subject constraints fail to detect it, see model 3 metrics in the Appendix), the distinction between *sembra che* and *sembrare*+infinitive seems to involve a more balanced mix of linguistic and extra-linguistic information, and overall, the detection of *sembra che* seems to be based slightly more on argument structure features. That said, it is safe to consider this constructional choice (or some part of it) to be strongly influenced by the properties of text genres, and this fact leads to some theoretical considerations for the study of alternations in CxG.

7.2. On including (features of) text genres in the representation of constructions

72 As a matter of fact, the relevance of extra-linguistic factors in the study of alternations has been highlighted by several studies. By “extra-linguistic” or “contextual” information here we refer to different aspects: real-world knowledge (Cappelle 2009); discourse environments (Rappaport Hovav 2014); stylistic and sociolinguistic variables (Lorenz & Levshina 2022). In some cases, such as ours, contextual factors have been proven to overtake purely linguistic ones (e.g., Lorenz & Levshina 2022). This could lead to a partial reconsideration of the assumption of No Synonymy between alternating constructions in CxG (Goldberg 1995), since in our analysis “extra-linguistic” factors can explain an important part of the variation, while linguistic ones seem to play a minor role. Even though different *sembrare*-constructions show indeed some differences in the prominence accorded to some elements, it could be argued that they can be used interchangeably in many contexts (also given the relative rarity of constrained linguistic environments). Thus, in our case, the No Synonymy should be embodied by differences in the degree of adequacy to different communicative situations more than by purely functional differences. Such considerations go in the direction of the readjustment of the Principle of No Synonymy as Principle of No Equivalence proposed in Leclercq & Morin (2023). Crucially, not only the name of the principle makes explicit that truth-conditional synonymy between two constructions is indeed possible, but in its formulation it also includes the possibility for two different forms to be distinct from the point of view of social meaning:

The Principle of No Equivalence: If two competing constructions differ in form (i.e. phonologically, morpho-syntactically or even orthographically), they must be semantically, pragmatically and/or *socially* distinct. (Leclercq & Morin 2023: 12, emphasis mine)

73 Socially meaningful differences pertain “recurrent elements of speaker background and socio-cultural context that come to be entrenched and conventionalised in socially-specific utterance types” (Leclercq & Morin 2023: 11), and include sociolinguistic variables (age, gender, etc.) and register information, compatibly with the contextual differences highlighted in our analysis. In such a scenario, it is reasonable to advocate for the inclusion of context-related information, in particular of text genres, in our

representations of the constructions. A problem, however, is that in of our approach we need to include properties which are not restricted to only one genre: that is, we cannot say that *mi sembra che* categorically appears in fiction prose or in dialogues, but that it is favoured instead by communicative situations that present some specific features. While it is impossible to adequately address the topic here, we can briefly suggest a possible solution to this problem.

74

An advantage of CxG framework is that it is possible to model text genres and text types as constructions: for instance, Hoffmann & Bergs (2018) consider text genres and types as meso- or macro-constructions that are “prototypical arrangement[s] of features that belong to this particular text type”. Such features can include genre-specific patterns and themes (e.g., a rhyme structure or *topoi*) that speakers abstract over the actual instances encountered (texts). Such a perspective makes possible to think of linguistic constructions (e.g., *mi sembra che*) as fillers for slots in highly abstract constructions (text genres), which have constraints on the elements appearing in their actual instances (textual productions). Thus, a linguistic construction can be more or less associated with a text genre, in the same way in which a lexeme can be associated or not with a slot in a linguistic construction. In our analysis such constraints are general properties, since we are not dealing with specific subgenres such as crime fiction, but with their “mother constructions” (e.g., fiction prose as a whole). For our purposes, we suggest that our representation of text genres includes some fundamental features that structure our experience of such communicative situations: the presence of interlocutors, their relationship with the speaker, if production and reception take place synchronically or not (cf. the criteria mentioned in Section 3.1), as well as other genre-specific features (e.g., need for conciseness, presence of dialogic sections). In turn, we can specify such features in constructions as preferred contextual environments in which said linguistic element is used: *mi sembra che* is favoured in contexts where an interlocutor is present, or in case there is less time to edit linguistic productions (synchronic nature of reception and production), and so on. In this way, we would not restrict linguistic constructions to be prototypical of only one text genre, and we can instead specify which conditions favour their use. An informal sketch of this proposal can be seen in Figure 6.

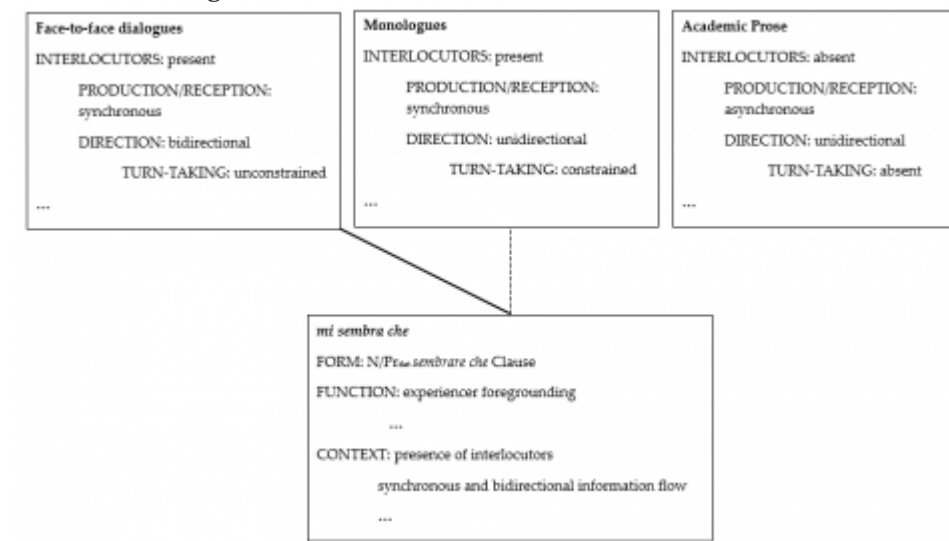


Figure 6. A model of the association between *mi sembra che* and different genre-constructions

75

Thus, in cases in which contextual information in a construction matches with text genre features, there will be a strong association¹¹, making it more probable that such a construction will appear in a text of that specific genre (*mi sembra che* and *Face-to-face dialogues* in Figure 6). Moreover, this proposal accounts also for the gradient nature of our approach: for instance, in monologues there are interlocutors, but there is less dialogic interaction (the information flow is unidirectional); thus, *mi sembra che* has a weaker association with *Monologues* (represented by the dotted link) but still can appear relatively frequently in that genre-construction, differently from genre-

constructions where there is no relationship with the interlocutors (e.g. *Academic Prose*).

- 76 While this is just an informal sketch, we believe that the proposed approach could be a fruitful direction to enrich our analysis of the relationship between constructional variability and contexts of use. Further research and formalization of text genres will hopefully shed light on the feasibility of this analysis.

8. Conclusions

- 77 Our study showed the relevance of extra-linguistic context in analyzing syntactic alternations. Furthermore, it highlighted the benefits of considering communicative situations (and thus, text genres) as a continuum and not as two prototypical opposing poles (as, for example, written and spoken language). While the prominence of 'external' factors in predicting speakers' choice in a syntactic alternation is not new, in the case of constructions with *sembrare*, linguistic constraints are only able to motivate a small amount of variation, while contexts of use are crucial to predict variability in speakers' choice.
- 78 Such results highlight that the speaker knows which text genres favour a given construction, and therefore this information is part of their representation of said construction. Thus, we believe it necessary to encode these values in the representation of the constructions even in cases such as ours (i.e., when we are dealing with "traditional", not genre-specific constructions). In cases of non genre-specific constructions, our proposal is to adopt a continuum-based approach to text genres (e.g., Biber 1988; Koch & Oesterreicher 2001; Voghera 2017), both in the analysis and in the representation of the contextual features of constructions. This approach has two advantages: it allows to conduct a finer grained analysis of the contexts of use, and it provides a solid criterion to understand the features that favour the distribution of a construction across different text genres (depending on the criteria according to which genres are ordered). Moreover, these criteria can be thought of as text genre properties in order to model their constructional status of environments for the use of linguistic patterns.
- 79 In conclusion, our results encourage to take more seriously into account extra-linguistic context in the analysis and the representation of constructions, and thus to use balanced and register-diversified corpora in our studies in order to achieve solid generalizations and capture variation in language use more faithfully.

Bibliographie

Achard, Michel. 2000. The distribution of French raising constructions. In Lisa J. Conathan, Jeff Good, Darya Kavitskaya, Alyssa B. Wulf & Alan C.L. Yu (eds.), *Proceedings of the 26th Annual Meeting of the Berkeley Linguistics Society, Berkeley, 18-21 February 2000*, 1-15. Ann Arbor: Sheridan Books.

Aijmer, Karin. 2009. *Seem* and evidentiality. *Functions of Language* 16(1). 63-88. DOI: 10.1075/fol.16.1.05aij

Auer, Peter & Stefan Pfänder. 2011. Constructions: Emergent or emerging? In Peter Auer & Stefan Pfänder (eds.), *Constructions: Emerging and Emergent*, 1-21. Berlin & Boston: De Gruyter. DOI: 10.1515/9783110229080.1

Bergs, Alexander. 2010. Expressions of futurity in contemporary English: A Construction Grammar perspective. *English Language and Linguistics* 14(2). 217-238. DOI: 10.1017/S1360674310000067

Berruto, Gaetano. 1985. Per una caratterizzazione del parlato: l'italiano parlato ha un'altra grammatica? In Günter Holtus & Edgar Radtke (eds.), *Gesprochenes Italienisch in Geschichte und Gegenwart*, 120-153. Tübingen: Gunter Narr.

Biber, Douglas. 1986a. On the investigation of spoken/written differences. *Studia Linguistica* 40. 1-21. DOI : 10.1111/j.1467-9582.1986.tb00759.x

- Biber, Douglas. 1986b. Spoken and Written Textual Dimensions in English: Resolving the Contradictory Findings. *Language* 62(2). 384-414. DOI: 10.2307/414678
- Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511621024
- Biber, Douglas. 2014. Using multi-dimensional analysis to explore cross-linguistic universals of register variation. *Languages in Contrast* 14(1). 7-34. DOI: 10.1075/lic.14.1.02bib
- Biber, Douglas & Conrad, Susan. 2009. *Register, genre, and style*. Cambridge-New York: Cambridge University Press. DOI: 10.1017/CBO9780511814358
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Pearson Longman.
- Blanche-Benveniste, Claire. 1997. La notion de variation syntaxique dans la langue parlée. *Langue Française* 115. 19-29.
- Calaresu, Emilia Maria. 2005. Quando lo scritto si finge parlato. La pressione del parlato sullo scritto e i generi scritti più esposti: il caso della narrativa. In Klaus Hölker, *Aspetti dell'italiano parlato*, 65-92. Münster: LIT Verlag.
- Calaresu, Emilia Maria & Palermo, Massimo. 2021. Iper testi e iperdiscorsi. Proposte di aggiornamento del modello di Koch e Oesterreicher alla luce dei testi nativi digitali. In Teresa Gruber, Klaus Gröbl & Thomas Scharinger (eds.), *Was bleibt von kommunikativer Nähe und Distanz? Mediale und konzeptionelle Aspekte sprachlicher Variation*, 82-113. Tübingen: Gunter Narr.
- Cappelle, Bert. 2009. Contextual cues for particle placement. Multiplicity, motivation, modeling. In Alexander Bergs & Gabriele Diewald (eds.), *Contexts and Constructions*, 145-192. Amsterdam: John Benjamins. DOI: 10.1075/cal.9.07cap
- Cerruti, Massimo. 2009. Premesse per uno studio della variazione di registro in italiano. *Rivista Italiana di Dialettologia. Lingue dialetti società* 33. 267-282. DOI: 10.1400/148739
- Chafe, Wallace. 1985. Linguistic Differences Produced by Differences between Speaking and Writing. In David R. Olson, Andrea Hildyard & Nancy Torrance (eds.), *Literacy, Language, and Learning*, 105-123. Cambridge: Cambridge University Press.
- Chafe, Wallace. 1986. Evidentiality in English conversation and academic writing. In Wallace Chafe & Johanna Nichols (eds.), *Evidentiality: The linguistic coding of epistemology*, 261-272. Norwood: Ablex.
- Chafe, Wallace & Deborah Tannen. 1987. The relation between written and spoken language. *Annual Review of Anthropology* 16. 383-407. DOI: 10.1146/annurev.an.16.100187.002123
- Cornillie, Bert. 2007. *Evidentiality and Epistemic Modality in Spanish (Semi-)Auxiliaries: A Cognitive-Functional Approach*. Berlin: Mouton De Gruyter. DOI: 10.1515/9783110204483
- Davies, Mark. 1997a. A Computer Corpus-Based Study of Subject Raising in Modern Portuguese. *Linguisticæ Investigationes* 21(2). 379-400. DOI: 10.1075/li.21.2.07dav
- Davies, Mark. 1997b. A Corpus-Based Analysis of Subject Raising in Modern Spanish. *Hispanic Linguistics* 9(1). 33-63.
- Fillmore, Charles J., Paul Kay & Mary C. O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language* 64(3). 501-538. DOI: 10.2307/414531
- Frascarelli, Mara. 2003. Topicalizzazione e ripresa clitica. Analisi sincronica, confronto diacronico e considerazioni tipologiche. In Teresa Poggi Salani & Nicoletta Maraschio (eds.), *Italia linguistica anno Mille. Italia linguistica anno Duemila: Atti del XXXIV Congresso internazionale di studi della Società di linguistica italiana (SLI), Firenze 19-21 ottobre 2000*, 1-16. Roma: Bulzoni Editore. DOI: 10.1400/28361
- Fried, Mirjam. 2009. Representing contextual factors in language change: Between frames and constructions. In Alexander Bergs & Gabriele Diewald (eds.), *Contexts and Constructions*, 63-94. Amsterdam: John Benjamins. DOI: 10.1075/cal.9.04fri
- Fried, Mirjam. & Jan-Ola Östman. 2005. Construction Grammar and spoken language: The case of pragmatic particles. *Journal of Pragmatics* 37(11). 1752-1778. DOI: 10.1016/j.pragma.2005.03.013
- Geka, Vassiliki, Sophia Marmaridou & Kiki Nikiforidou. 2020. Dialogic constructions and discourse units: The case of *think again*. *Review of Cognitive Linguistics* 18(2). 480-518. DOI: 10.1075/rcl.00068.gek.
- Gisborne, Nikolas & Jasper Holmes. 2007. A History of English Evidential Verbs of Appearance. *English Language and Linguistics* 11(1). 1-29. DOI: 10.1017/S1360674306002097
- Goldberg, Adele E. 1995. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Grimm, Scott. 2010. An Empirical View on Raising to Subject. *ZAS Papers in Linguistics* 52. 83-109. DOI: 10.21248/zaspil.52.2010.384

- Hilpert, Martin. 2016. Change in modal meanings: Another look at the shifting collocates of *may*. *Constructions and Frames* 8(1). 66-85. DOI: 10.1075/cf.8.1.05hil.
- Hilpert, Martin. 2019. *Construction Grammar and its application to English*. 2nd edition. Edinburgh: Edinburgh University Press.
- Hilpert, Martin & Holger Diessel. 2017. Entrenchment in Construction Grammar. In Hans-Jörg Schmid (ed.), *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*, 57-74. Berlin/Boston: De Gruyter. DOI: 10.1515/9783110341423-004.
- Hoffman, Thomas & Alexander Bergs. 2018. A Construction Grammar Approach to Genre. *CogniTextes* 18. DOI: 10.4000/cognitextes.1032.
- Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* 15(3). 651-674. DOI: 10.1198/106186006X133933.
- Hothorn, Torsten & Achim Zeileis. 2015. *partykit*: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research* 16(118). 3905-3909. URL: <https://www.jmlr.org/papers/v16/hothorn15a.html>
- Iacobini, Claudio. 2008. Presenza e uso dei verbi sintagmatici nel parlato dell'italiano. In Monica Cini (ed.), *I verbi sintagmatici in italiano e nelle varietà dialettali. Stato dell'arte e prospettive di ricerca- Atti delle giornate di studio Torino, 19.-20. febbraio 2007*, 103-119. Frankfurt am Main: Peter Lang. DOI: 10.3726/978-3-653-01094-7
- Kayne, Richard. 1981. On certain differences between French and English. *Linguistic Inquiry* 12(3). 349-371.
- Kirk, John M. 1997. Subordinate Clauses in English. *Journal of English Linguistics* 25(4). 349-364. DOI: 10.1177/007542429702500409
- Koch, Peter & Wulf Oesterreicher. 2001. Gesprochene Sprache und geschriebene Sprache. Langage parlé et langage écrit. In Günter Holtus, Michael Metzeltin & Christian Schmitt (eds.), *Lexicon der Romanischen Linguistik* (LRL). Vol I.2, 584-627. Tübingen: Niemeyer.
- Kratschmer, Alexandra R. 2005. Raising- og Small Clause-konstruktioner med italiensk *sembrare* /*parere*: polyfoni og evidentialitet. In Henning Nølle, Rita Therkelsen & Nina Møller Andersen (eds.), *Sproglig polyfoni: Arbejdsrapporter 3: Juli 2005*, 101-137. Roskilde: Roskilde Universitetscenter, Inst. f. Sprog og Kultur.
- Kratschmer, Alexandra R. 2006. Che te ne sembra? Semantica e pragmatica delle costruzioni italiane con *sembrare*/*parere*. In Michel Olsen & Erik H. Swiatek (eds.), *XVI Congreso de Romanistas Escandinavos/XVI Congrès des Romanistes Scandinaves/XVI Congresso dei Romanisti Scandinavi/XVI Congresso dos Romanistas Escandinavos*. Roskilde: Roskilde Universitetscenter, Inst. f. Sprog og Kultur.
- Lacheret-Dujour, Anne, Sylvain Kahane & Paola Pietrandrea (eds.). 2019. *Rhapsodie. A prosodic and syntactic treebank for spoken French*. Amsterdam / Philadelphia: John Benjamins. DOI: 10.1075/scl.89
- Langacker, Ronald W. 1995. Raising and Transparency. *Language* 71(1). 1-62. DOI: 10.2307/415962
- Larsson, Tove & Kataari, Henrik. 2020. Syntactic complexity across registers: Investigating (in)formality in second-language writing. *Journal of English for Academic Purposes* 45. DOI: 10.1016/j.jeap.2020.100850
- Lavinio, Cristina. 1990. *Teoria e didattica dei testi*. Firenze: La Nuova Italia.
- Lavinio, Cristina. 2000. Tipi testuali e processi cognitivi. In Fabio Camponovo & Alessandra Moretti (eds.), *Didattica ed educazione linguistica*, 125-144, La Nuova Italia: Firenze. http://giscel.it/wp-content/uploads/documenti/pubblicazioni/2000/2000_QuaderniNS1.pdf
- Leclercq, Benoît & Cameron Morin. 2023. No equivalence: A new principle of no synonymy. *Constructions* 15(1). DOI : 10.24338/cons-535
- Levshina, Natalia. 2021. Conditional inference trees and random forests. In Magali Paquot & Stefan Th. Gries (eds.), *Practical Handbook of Corpus Linguistics*, 611-643. New York: Springer. DOI: 10.1007/978-3-030-46216-1_25
- Levshina, Natalia & David Lorenz. 2022. Communicative efficiency and the Principle of No Synonymy: Predictability effects and the variation of *want to* and *wanna*. *Language and Cognition* 14(2). 249-274. DOI: 10.1017/langcog.2022.7
- Linell, Per. 1982. *The Written Language Bias in Linguistics*. Linköping: University of Linköping.
- Linell, Per. 2009. Grammatical constructions in dialogue. In Alexander Bergs & Gabriele Diewald (eds.), *Contexts and Constructions*, 97-110. Amsterdam: John Benjamins. DOI: 10.1075/cal.9.05lin
- Masini, Francesca, Caterina Mauri & Paola Pietrandrea. 2018. List constructions: Towards a unified account. *Italian Journal of Linguistics* 30. 49-94. DOI: 10.26346/1120-2726-116

- Matsumoto, Yoshiko. 2010. Interactional Frames and Grammatical Description. *Constructions and Frames* 2(2). 135-157. DOI : 10.1075/cf.2.2.01mat
- Miecznikowski, Johanna. 2020. At the juncture between evidentiality and argumentation: evidential verb complementation. *Journal of Argumentation in Context*, 9(1). 42-68. DOI: 10.1075/jaic.00007.mie
- Miecznikowski, Johanna & Elena Musi. 2015. Verbs of appearance and argument schemes: Italian *sembrare* as an argumentative indicator. In Frans van Eemeren & Bart Garssen (eds.), *Reflections on Theoretical Issues in Argumentation Theory*, 259-278. Amsterdam: Springer. DOI : 10.1007/978-3-319-21103-9_19
- Montague, Richard M. 1970. Pragmatics and intensional logic. *Synthese* 22(1). 68-94.
- Moro, Andrea. 1992. *I predicati nominali e la struttura della frase*. Padova: Unipress.
- Mortelmans, Tanja. 2017. Seem-type verbs in Dutch and German. In Juana I. Marín-Arrese, Gerda Haßler & Marta Carretero (eds.), *Evidentiality Revisited: Cognitive grammar, functional and discourse-pragmatic perspectives*, 123-148. Amsterdam-Philadelphia: John Benjamins. DOI: 10.1075/pbns.271.06mor
- Musi, Elena. 2015. *Dalle apparenze alle inferenze: i predicati sembrare e apparire come indicatori argomentativi*. Lugano: Università della Svizzera Italiana dissertation. URL: <https://susi.usi.ch/usi/documents/318627>
- Musi, Elena. 2016. Sembrare tra semantica e sintassi: proposta di un'annotazione multilivello. In Annibale Elia, Claudio Iacobini & Miriam Voghera (eds.), *Atti del XLVII congresso internazionale di studi della società di linguistica italiana (SLI) – Fisciano-Salerno 26-28 settembre 2013*, 159-176. Roma: Bulzoni.
- Nencioni, Giovanni. 1976. Parlato-parlato, parlato-scritto, parlato-recitato. *Strumenti critici* LX. 1-56.
- Neset, Tore & Laura Janda. 2023. A network of allostructions: quantified subject constructions in Russian. *Cognitive Linguistics* 34(1). 67-97. DOI : 10.1515/cog-2021-0117
- Nikiforidou, Kiki. 2010. Viewpoint and construction grammar: The case of past + now. *Language and Literature* 19(3). 265-284. DOI: 10.1177/0963947010370253
- Nikiforidou, Kiki. 2016. 'Genre knowledge' in a constructional framework: Lexis, grammar and perspective in folk tales. In Ninke Stukker, Wilbert Spooren & Gerard Steen (eds.), *Genre in Language, Discourse and Cognition*, 331-360. Berlin & Boston: De Gruyter Mouton. DOI: 10.1515/9783110469639-014
- Nikiforidou, Kiki 2021. Grammatical variability and the grammar of genre: Constructions, conventionality, and motivation in 'stage directions'. *Journal of Pragmatics* 173. 189-199. DOI: 10.1016/j.pragma.2020.12.012
- Nissim, Malvina, Shipra Dingare, Jean Carletta & Mark Steedman. 2004. An Annotation Scheme for Information Status in Dialogue. In Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (eds.), *Proceedings of the 4th Conference on Language Resources and Evaluation (LREC'04)*. <https://aclanthology.org/L04-1402/>
- Ochs, Elinor. 1979. Planned and Unplanned Discourse. In Talmy Givón (ed.), *Discourse and Syntax*, 51-80. New York: Academic Press. DOI: 10.1163/9789004368897_004
- Oesterreicher, Wulf. 1997. Types of Orality in Text. In Egbert J. Bakker & Ahuvia Kahane (eds.), *Written Voices, Spoken Signs: Tradition, Performance, and the Epic Text*, 190-214. Cambridge, MA/London: Harvard University Press. DOI: 10.4159/9780674020467-010
- Östman, Jan-Ola. 2005. Construction Discourse. A prolegomenon. In Jan-Ola Östman & Mirjam Fried (eds.), *Construction Grammars: Cognitive grounding and theoretical extensions*, 121-144. Amsterdam: John Benjamins. DOI: 10.1075/cal.3.06ost
- Pietrandrea, Paola. 2018. Epistemic Constructions at work. A Corpus Study on Italian Dialogues. *Journal of Pragmatics* 128. 171-191. DOI: 10.1016/j.pragma.2017.10.006
- Pisciotta, Flavio. 2023. The alternation between “raising” and impersonal constructions with *sembrare*: a usage-based approach. In Cristiana Cervini & Gloria Gagliardi (eds.), *CLUB Working Papers in Linguistics. Volume 7*, 95–122. Bologna: CLUB – Circolo Linguistico dell'Università di Bologna.
- Pisciotta, Flavio & Francesca Masini. accepted. *A paradigm of psych-predicates: unraveling the constructional competition between light verbs constructions and derived verbs in Italian*. [Manuscript accepted for publication].
- Pistolessi, Elena. 2015. Diamesia: la nascita di una dimensione. In Elena Pistolessi, Rosa Pugliese, Barbara Gili Fivela (eds.), *Parole, gesti, interpretazioni. Studi linguistici per Carla Bazzanella*, 27-56. Roma: Aracne. DOI: 10.4399/97888548840762
- Polinsky M. 2013. Raising and control. In Marcel den Dikken (ed.), *The Cambridge Handbook of Generative Syntax*, 577-606. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511804571.021

- Postal, Paul M. 1970. On the surface verb 'remind'. *Linguistic inquiry* 1(1). 37-120.
- Postal, Paul M. 1974. *On Raising: One rule of English grammar and its theoretical implications*. Cambridge, MA: MIT Press.
- Raible, Wolfgang. 2019. Variation in Language: How to Characterise Types of Texts and Communication Strategies between Orality and Scripturality. Answers given by Koch/Oesterreicher and by Biber. *International Journal of Language and Linguistics* 6(2). 157-174. DOI : 10.30845/ijll.v6n2p19
- Rappaport Hovav, Malka. 2014. Lexical content and context: The causative alternation in English revisited. *Lingua* 141. 8-29. DOI : 10.1016/j.lingua.2013.09.006
- Rooryck, Johan. 1997. On the Interaction Between Raising and Focus in Sentential Complementation. *Studia Linguistica* 51(1). 1-49. DOI: 10.1111/1467-9582.00016
- Rosenbaum, Peter. 1967. *The grammar of English predicate complement constructions*. Cambridge, MA: MIT Press.
- Ruppenhofer, Josef & Laura A. Michaelis. 2010. A constructional account of genre-based argument omissions. *Constructions and Frames* 2(2). 158-184. DOI: 10.1075/cf.2.2.02rup
- Salvi, Giampaolo. 1991. I complementi predicativi. In Lorenzo Renzi, Giampaolo Salvi & Anna Cardinaletti (eds.), *Grande Grammatica italiana di consultazione. Volume 2: I sintagmi verbale, aggettivale, avverbiale. La subordinazione*, 191-226. Bologna: Il Mulino.
- Shahrokny-Prehn, Arian & Silke Höche. 2011. Rising through the registers – A corpus-based account of the stylistic constraints on Light Verb Constructions. *Corpus* 10. DOI: 10.4000/corpus.2110
- Szmrecsanyi, Benedikt & Lars Hinrichs. 2008. Probabilistic determinants of genitive variation in spoken and written English: A multivariate comparison across time, space, and genres. In Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta & Minna Korhonen (eds.), *The Dynamics of Linguistic Variation: Corpus evidence on English past and present*. 291–309. Amsterdam: John Benjamins. DOI: 10.1075/silv.2.22szm
- Szmrecsanyi, Benedikt, Jason Grafmiller, Benedikt Heller & Melanie Röthlisberger. 2016. Around the world in three alternations. Modeling syntactic variation in varieties of English. *English World-Wide* 37(2). 109-137. DOI: 10.1075/eww.37.2.01szm
- Thompson, Susan. 1994. Aspects of Cohesion in Monologue. *Applied Linguistics* 15(1). 58–75. DOI: 10.1093/applin/15.1.58
- Torrego, Esther. 1996. Experiencers and raising verbs. In Robert Freidin (ed.), *Current Issues in Comparative Grammar*, 101-120. Dordrecht & Boston: Kluwer Academic Publishers. DOI: 10.1007/978-94-009-0135-3_5
- Vliegen, M. 2011. Scheinbar identisch: Niederländisch *schijnen*, deutsch *scheinen*. In Wilfried Kürschner, Reinhard Rapp, Jürg Strässler, Maurice Vliegen & Heirich Weber (eds.), *Neue linguistische Perspektiven*, 231-244. Frankfurt am Main: Peter Lang.
- Voghera, Miriam. 1992. *Sintassi e intonazione nell'italiano parlato*. Bologna: Il Mulino.
- Voghera, Miriam. 2017. *Dal parlato alla grammatica. Costruzione e forma dei testi spontanei*. Roma: Carocci.
- Voghera, Miriam. 2022. Scritto-parlato e altri modi nell'educazione. *Italiano LinguaDue* 14(2). 1-18. DOI: 10.54103/2037-3597/19646
- Voghera, Miriam, Claudio Iacobini, Renata Savy, Francesco Cutugno, Aurelio De Rosa, Iolanda Alfano. 2014. VoLIP: a searchable Italian spoken corpus. In Ludmila Veselovská & Markéta Janebová (eds.), *Complex Visibles Out There. Proceedings of the Olomouc Linguistics Colloquium 2014, Language Use and Linguistic Structure*, 627-640. Olomouc: Palacký University Olomouc.
- CORIS = Rossini Favretti, Rema, Fabio Tamburini & Cristiana De Santis. 2002. CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In Andrew Wilson, Paul Rayson & Tony McEnery (eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, 27-38. München: LINCOM-Europa. Available at: <http://corpora.dslo.unibo.it/TCORIS/>
- KIParla = Mauri, Caterina, Silvia Ballarè, Eugenio Gorla, Massimo Cerruti & Francesco Suriano. 2019. KIParla corpus: a new resource for spoken Italian. In Raffaella Bernardi, Roberto Navigli & Giovanni Semeraro (eds.), *Proceedings of the Sixth Italian Conference on Computational Linguistics (CLIC-it 2019)*. CEUR Workshop Proceedings. URL: <http://ceur-ws.org/Vol2481/paper45.pdf> Available at: <https://kiparla.it/search>
- LIP = De Mauro, Tullio, Federico Mancini, Massimo Vedovelli & Miriam Voghera. 1993. *Lessico di frequenza dell'italiano parlato*. Milano: Etaslibri. Available at: <https://www.volip.it/>

Notes

1 Henceforth we will refer to (1a) as *sembra che* and to (1b) as *sembrare*+infinitive.

2 I use the label “text genres” to refer both to written genres (fiction prose, press, etc.) and to spoken communicative situations (e.g., dialogues, monologues, etc.). It is important to note that “text genres” here overlaps with the notion of register in corpus linguistics (see Biber & Conrad 2009).

3 The difference in the distribution of the constructions is highly significant : the chi-square test yields a p-value < 0.0001. The analysis of the standardised residuals shows that the cells that contribute most to the significance of the test (i.e., containing standardized residuals higher than +2 and lower than -2) are the *mi sembra che* and *sembrare*+infinitive ones :

sembra che mi sembra che sembrare+infinitive

Written	-1.1	-9.1	9.7
Spoken	1.1	9.1	-9.7

4 In Table 1 there is no distinction between *sembrare*+infinitive with and without experiencer, since only two constructions with overt experiencer were found.

5 Interestingly, if we look at studies carried out on other European languages, there seems to be a discrepancy between Romance and Germanic languages: as a matter of fact, in English (Biber et al. 1999; Grimm 2010), German and Dutch (Vliegen 2011; Mortelmans 2017) the infinitive construction with SEEM-verbs is overwhelmingly more used than the impersonal one in both writing and speech. However, this aspect needs to be further investigated by gathering data from more languages and in a comparable fashion.

6 This annotation scheme describes the relationship between the subject NP and its antecedent. As an example, Old_identity applies to NPs that are simply coreferential with their antecedent, Old_generic applies to pronouns used for generic reference, etc. For a full description of the annotation scheme, cf. Nissim et al. 2004 or the appendix in Pisciotta (2023).

7 The R code and output is available in the Appendix. I thank the anonymous reviewer for suggesting the reduction of the split size, and for the precious comments and suggestions on the methodology.

8 On the contrary, *sembra che* can express a judgement not necessarily endorsed by the speaker :




[...] *io sembra che fumo ma non fumo* ‘me, it seems that I smoke, but I don’t’ (LIP, NE5)

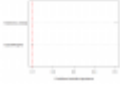


9 Cf. also Berruto’s (1985) notion of egocentrism as a central feature of spoken language.

10 This behavior has been noticed in other studies on alternations, and discussed by distinguishing between cue reliability and cue availability (cf. Nessel & Janda 2023).

11 In Figure 6, we model such relationship similarly to filler-slot relations (Hilpert & Diessel 2017), in order to maintain a parallelism with such kind of links (see above). However, by this we do not imply at this stage any homology between filler-slot relations and “genre-construction” relations.

Table des illustrations

	Légende	Figure 1. The dialogicity and planning scale of text genres described in Voghera (2017)
	URL	http://journals.openedition.org/cognitextes/docannexe/image/3017/img-1.png
	Fichier	image/png, 28k
	Légende	Figure 2. Proportions of the constructions for each text genre, plotted against Voghera’s (2017) scale.
	URL	http://journals.openedition.org/cognitextes/docannexe/image/3017/img-2.png
	Fichier	image/png, 39k
	Légende	Figure 3. Conditional Inference Tree distinguishing between the three sembrare-constructions

	URL	http://journals.openedition.org/cognitextes/docannexe/image/3017/img-3.jpg
	Fichier	image/jpeg, 1,1M
	Légende	Figure 4. Relative importance of the variables for distinguishing between the three sembrare-constructions
	URL	http://journals.openedition.org/cognitextes/docannexe/image/3017/img-4.png
	Fichier	image/png, 11k
	Légende	Figure 5. Relative importance of the variables for distinguishing between the three sembrare-constructions (TextGenre_Ordinal removed).
	URL	http://journals.openedition.org/cognitextes/docannexe/image/3017/img-5.png
	Fichier	image/png, 10k
	Légende	Figure 6. A model of the association between mi sembra che and different genre-constructions
	URL	http://journals.openedition.org/cognitextes/docannexe/image/3017/img-6.png
	Fichier	image/png, 31k

Pour citer cet article

Référence électronique

Flavio Pisciotta, « When constructional choice is a matter of context: *sembrare*-constructions across a continuum of text genres », *CogniTextes* [En ligne], Volume 25 | 2024, mis en ligne le 23 novembre 2024, consulté le 26 novembre 2024. URL : <http://journals.openedition.org/cognitextes/3017> ; DOI : <https://doi.org/10.4000/12rcy>

Auteur

Flavio Pisciotta
University of Salerno

Droits d'auteur



Le texte seul est utilisable sous licence CC BY-NC-ND 4.0. Les autres éléments (illustrations, fichiers annexes importés) sont « Tous droits réservés », sauf mention contraire.