
The alternation between “raising” and impersonal constructions with *sembrare*: a usage-based approach

Flavio Pisciotta

Università di Salerno

fpisciotta@unisa.it

Abstract

Traditionally, scholars have claimed that the raising construction with *sembrare* ‘seem’ (*sembrare*+infinitive) is truth-conditionally equivalent to the impersonal construction *sembra che* ‘it seems that’. In contrast with this point of view, and following a Construction Grammar approach, we hypothesize that this alternation is motivated by differences in the functional properties of the alternating constructions. To test this hypothesis, we focused on selection properties, with the view of assessing whether any subject could be a good candidate to appear in the constructions under exam. We extracted occurrences of the two constructions from written and spoken Italian corpora, and coded their subjects for several formal and functional variables (length, animacy, accessibility, and type of anaphoric link). Our analysis showed that there are, in fact, differences in the selection of the subjects, and that such differences reveal functional discrepancies between the two constructions. Furthermore, we found that the presence of the experiencer can constrain the semantic contexts in which *sembra che* appears. These results suggest the relevance of the concept of “allostruction” to account for the behaviour of alternating constructions.

1. Introduction

The Italian verb *sembrare* ‘seem’ has attracted many scholars working on syntax and semantics. In particular, the syntactic constructions in which personal *sembrare* takes an infinitive clause are characterized by the so-called subject-to-subject raising (which we will refer to simply as “raising”) (1a).¹ *Sembrare* and similar verbs are thus defined as raising verbs, in contrast to control verbs (like *volere* ‘want’), which select subject control constructions (1b):

¹ This paper is based on (and updates) part of the analysis carried out in my Master's thesis *Le costruzioni impersonali e a sollevamento con sembrare in italiano scritto e parlato: uno studio costruzionista e usage-based*. I thank prof. Francesca Masini and prof. Chiara Gianollo who, respectively, supervised and co-supervised the thesis. I also thank the anonymous reviewer for the insightful comments and suggestions.

- (1) a. *Gianni sembra amare Piero.*
 ‘Gianni seems to love Piero.’
 b. *Gianni vuole sposare Piero.*
 ‘Gianni wants to marry Piero.’

The concept of raising originates from the generativist tradition and refers to a movement-based account of the phenomenon, according to which the subject selected by the subordinate verb “raises” to the main clause and becomes the subject of the main verb² (Rosenbaum 1967; Postal 1974; Polinsky 2013). We will not get into the details of the diagnostics used to distinguish raising verbs from subject control verbs (cfr. Skytte & Salvi 1991: 543-544; Davies & Dubinsky 2004: 4-8). However, it is worth noting that the main difference between the constructions in (1a) and (1b) is that in (1b) *Gianni* is the semantic subject of both the verbs *vole* ‘wants’ and *sposare* ‘to marry’. Conversely, in (1a) the NP *Gianni* is selected as the subject by the subordinate verb *amare* ‘to love’, and not by *sembrare*. For this reason, the raising construction with *sembrare* has been traditionally considered truth-conditionally equivalent to the construction with impersonal *sembrare* (2) (Davies & Dubinsky 2004: 4), also in non-transformational approaches (cfr. Bach 1977; Dowty 1978).

- (2) *Sembra che Gianni ami Piero.*
 ‘It seems that Gianni loves Piero.’

Both in (1a) and in (2), *Gianni* is selected by the subordinate verb, but the two sentences result from different transformations: in (2), *Gianni* does not raise to the main clause, and thus constitutes the subject selected by the subordinate verb not only semantically but also syntactically.

Nonetheless, some scholars from generative approaches have noticed meaning differences between the constructions in (1a) and (2). Postal (1974: 357) and Salvi (1991: 218-219) argue that the raising construction presupposes a context in which the speaker directly perceives the *stimulus* represented by the raised subject. Partee (1971: 17-18) also suggests that impersonal and raising constructions show some subtle meaning differences: in the former, the speaker’s point of view is focused on the whole event described by the subordinate, while in the latter it is focused only on the raised subject. Anyway, these proposals hold that the differences found are not part of the core meaning of these constructions (to be intended as their logical form). From this perspective, both (1a) and (2) are derived from the same deep structure and preserve the same truth conditions. This position raises questions about what aspects are relevant to establish if two constructions are synonymous or not (Mack 2010: 155). Furthermore, it does not help understanding the reasons that underlie the speaker’s choice of one of the two constructions, implicitly hypothesizing a free variation between the two.

While this could be the case, scholars working in usage-based frameworks, e.g., Construction Grammar, found that a number of syntactic alternations are motivated by different functions or properties associated with alternating formal variants (Gries 1999; 2003; Stefanowitsch 2003; Perek 2015; De Vaere et al. 2018). These analyses generally assume the so-called “Principle of No Synonymy” (cfr. Bolinger 1968). In particular,

² For a comprehensive overview of the generative accounts of raising, see Davies & Dubinsky (2004).

Goldberg (1995: 67) adopts this principle when describing the relation among constructions in the speaker's repertoire:

If two constructions are syntactically distinct, they must be semantically or pragmatically distinct. [...] Pragmatic aspects of constructions involve particulars of information structure, including topic and focus, and additionally stylistic aspects of the construction such as register.

Keeping in mind that a dogmatic approach to this principle would be empirically unmotivated³, we want to test if it applies to the *sembrare* alternation as well. Our hypothesis is that the constructions in (1a) and (2) differ with respect to their functional properties. Specifically, we want to investigate this aspect by looking for possible constraints or preferences in the selection of the element that fills the slot of the “raised” or “unraised” subject: potential differences between the two alternants would suggest that there is a difference between the semantics of the two constructions.

2. Previous studies on SEEM constructions

In this section, we illustrate previous studies on raising and impersonal constructions with SEEM verbs (henceforth, SEEM constructions). These constructions are renowned for being evidential and epistemic strategies (Section 2.1) and have been mainly studied with regard to this functional aspect. Nonetheless, there are also studies on the selection of the subject in this syntactic alternation. These studies have been generally carried out from a cognitive and/or corpus-based perspective and on languages different from Italian. We present them in Section 2.2 as a starting point for our case study.

2.1 *Sembrare constructions as epistential strategies*

In Italian and other European languages, SEEM constructions are used to convey evidential and epistemic meanings (Cornillie 2007; Gisborne & Holmes 2007; Aijmer 2009; Musi 2015). Although the present analysis will not focus on evidentiality and epistemic modality, they constitute a central aspect in the function of the constructions with *sembrare*, and thus it is worth introducing these concepts briefly.

We can define evidentiality as the category that marks the speaker's information source (Aikhenvald 2004). Information sources can be either direct (e.g., speaker's direct perception) or indirect (e.g., inferences, reports from a third person). This is a relevant distinction in our case, because in the literature it has been noted that there is often an overlap or, at least, a relationship between the marking of indirect sources and the expression of epistemic modality, i.e., the speaker's (un)certainly about the likelihood of the uttered proposition (Nuyts 2001: 21-22).

³ While the validity of the Principle is generally accepted by scholars working in Construction Grammar, some aspects have been subject to debate: e.g., the role of free variation between alternating constructions (Uhrig 2015), or the level of abstraction at which the principle applies (Laporte et al. 2021) (for a recent discussion on the matter, see Leclercq & Morin (2023)). As for our perspective, we want our approach to be as much empirical as possible, and thus we do not exclude the possibility that the constructions with *sembrare* are used in free variation.

This is the case with verbs denoting appearance, including *sembrare*. Because of its semantics, even when *sembrare* marks that the speaker is perceiving something, it also signals that there is a cognitive elaboration of the *perceptum*, in order to reconstruct the state of affairs (3). Thus, constructions with *sembrare* always express indirect evidentiality.

- (3) *Lei lo tiene in braccio sorridente, lui sembra dormire.*
 ‘She is holding him in her arms, he smiles and seems to be sleeping.’ (CORIS⁴)

Furthermore, the ‘appearance’ meaning always implies a degree of uncertainty: “what appears [...] is not necessarily claimed to be true” (Squartini 2018: 281). Hence, the constructions with *sembrare* have been defined as “epistentials”, that is, strategies that show a syncretism of evidential and epistemic values (Musi & Rocci 2017).

Research has shown differences in the meanings of *sembrare*-constructions at the epistential level. In Italian, the raising construction (3) has been claimed to encode only inferential evidentiality, while the impersonal can also mark that the information has been reported to the speaker or comes from hearsay (4) (Kratschmer 2006; Musi 2015):

- (4) *Sembra che i furti siano da imputare ad una banda [...].*
 ‘It seems that the thefts are to be blamed on a gang [...].’ (CORIS)

Moreover, some studies have identified a third construction: the impersonal construction with overt experiencer (5).

- (5) *Mi sembra che sia il caso di aiutarla [...].*
 ‘It seems to me that we should help her.’ (CORIS)

This construction has been said to encode mainly the subjective point of view of the speaker, thus emphasizing the epistemic judgement on the state of affairs (Squartini 2009: 220; Musi 2015: 327-328). Given this, we will consider the construction with overt experiencer as a distinct one, even though most research on the alternation between SEEM constructions does not treat it as such (e.g., Davies 1997; Achard 2000; Ruppenhofer 2004; Grimm 2010).

2.2 Discourse properties of raised subjects: between prominence and activation status

As mentioned in Section 1, research in formal syntax has generally assumed a truth-conditional equivalence between the two alternant constructions with SEEM verbs. In contrast to this view, cognitive approaches have stressed that the formal differences between the two constructions reflect a difference in the construal of the same situation (i.e., the way in which it is presented by the speaker) (Newman 1981; Langacker 1995).

Langacker (1995) claims that raised subjects function as reference points for the conceptualizer: they constitute a privileged “point of mental access” to the propositional content. For instance, in (6a) *Don* is the most salient participant in the process of leaving and has a more prominent role in the evaluation of the probability that he will leave (e.g.,

⁴ <https://corpora.ficlit.unibo.it/TCORIS/>

he is a more volitional participant) than in (6b):

- (6) a. *Don is likely to leave.* (Langacker 1995: 21)
b. *It is likely that Don will leave.*

Langacker defines this prominent element as “a sort of local topic” (1995: 38). Thus, the raised subject in (6a) is more topical than the unraised one in (6b).

Other accounts of these constructions have referred to topicality as a crucial difference between the two alternants (cfr. Croft 2001; Givón 2001; Barron 2001). This holds also for languages different from English: in her study on Italian, Musi finds that raised subjects are more likely to be more definite, referential and topical than unraised ones (2015: 423).

Nonetheless, the interpretation of the concept of topicality in these studies is not always unproblematic. In his study on the raising/impersonal alternation in French, Achard (2000) provides support for Langacker’s interpretation of raising. He operationalizes topicality through the notion of information status: an element is more topical if it has an active status in discourse. His findings show that raised subjects are more often given (i.e., discourse-old) than unraised ones. However, his reading of topicality, common to Musi’s analysis of Italian, can be misleading: research has shown that topical elements are often not discourse-old and that the overlap between these two notions (givenness and topicality) is complex and at least imperfect (Mack 2010: 51-52).

Furthermore, the results of these studies are not always coherent. Ruppenhofer (2004), adopting a Construction Grammar approach, investigates this alternation in English to test Langacker’s and Achard’s findings. The only significant effect he finds concerns the subject length: in his spoken data, raising constructions with *seem* select shorter subjects than the impersonal counterparts (Ruppenhofer 2004: 238). This could hint at the fact that raised subjects are more easily identifiable and more topical, but his results show that there is no effect of givenness and topicality on the choice of the construction. We could suppose that the difference between Achard’s and Ruppenhofer’s results might be due to language-specific differences, but another study on English, carried out by Grimm (2010), finds that raised subjects are significantly more discourse-old.

The only possible explanation for the discrepancies we find in the literature is that the criteria for the selection of the analyzed samples may be different or that there are methodological divergences in the annotation of givenness and topicality of the subject NPs. This last aspect is probably the result of the above-mentioned lack of unanimity on the definition of these two notions.

Summing up, while an effect of information status (i.e., givenness) and/or informational relations (e.g., topicality) on the production of raising constructions is plausible, the problem is far from being solved and will be tackled only partly in this article (Section 5.1).

3. Case study

3.1 Hypothesis: the constructions with *sembrare* as *allostructions*

Our study will be carried out within the framework of Construction Grammar (Fillmore et al. 1988; Goldberg 1995; Hoffmann & Trousdale 2013). According to this approach, syntactic structures should be treated as (more or less abstract) templates that include a

form and a function, namely as constructions. The latter are therefore form-meaning pairings with their own idiosyncratic properties, which cannot be deduced from the parts or derived from other constructions (Goldberg 1995: 4). Schematic (i.e., abstract) or partially schematic constructions have empty slots that can be filled only by certain elements, based on constraints specific to each construction.

Given that we will not follow a transformational approach, we will refer to the constructions under examination by their form: *sembrare*+infinitive (the raising construction), *sembra che* (the impersonal construction), *mi sembra che* (the impersonal construction with overt experiencer). Our main hypothesis is that *sembrare*+infinitive and *sembra che* do not behave uniformly with regard to subject selection: different preferences or constraints on subject selection would suggest the existence in the speaker's linguistic knowledge of two distinct objects, each endowed with its own properties, that cannot be accounted for by synchronic transformations.

However, the absence of a common deep structure between *sembrare*+infinitive and *sembra che* does not exclude a relationship between them (Cappelle 2006: 12). Construction Grammar posits that constructions are interconnected in a network-like fashion. The links between constructions may specify some kind of relationship between them (Lakoff 1987; Goldberg 1995): e.g., a semi-schematic construction could inherit some of its features from a more schematic construction. In our case, we hypothesize that *sembrare*+infinitive and *sembra che* are “allostructions”, i.e., constructions connected by a horizontal link (Cappelle 2006). Allostructional links connect two or more alternating constructions that show a similarity in their meaning, but that do not inherit this common meaning from a ‘mother’ node, since their formal divergence would make it impossible to postulate such a common ‘mother’ construction. Such a formalization captures that the speaker presumably knows that two formally unrelated constructions can be used, in some contexts, as alternating constructions, or as paraphrases of one another.

Some scholars formalize allostructions as alternating (sub)variants of an abstract, formally underspecified construction, the so-called “constructeme” (even though there is debate on the necessity of postulating its existence: cfr. Audring 2019 and Ungerer, forthcoming). The presence of a constructeme further helps to capture the fact that the speaker makes an abstract generalization on the two variants. In Figure 1, the ditransitive (e.g., *Paul sent Mary a letter*) and the to-dative constructions (e.g., *Paul sent a letter to Mary*) both inherit some functional and formal features from the dative constructeme, that encodes the overlapping information contained in the two allostructions. Nonetheless, they are distinct constructions, associated with – and motivated by – different discourse and pragmatic properties (Cappelle 2006; Perek 2015).

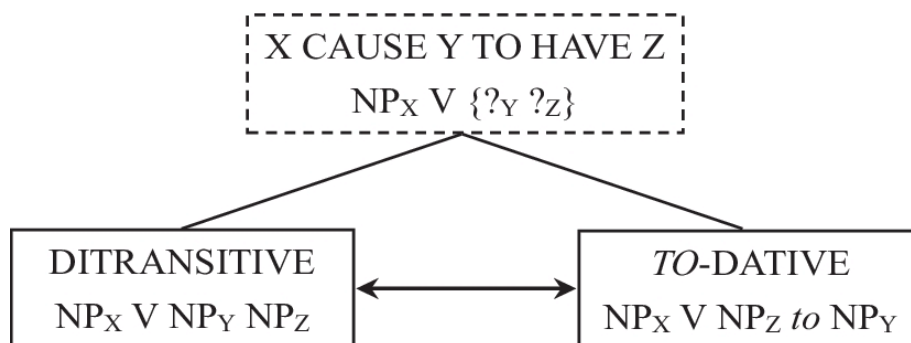


Figure 1. The dative constructeme and its two allostructions (Perek 2015: 156).

We argue that the relationship between *sembrare*+infinitive and *sembra che* can be fruitfully described by such a model.

Given that research has found that the overt expression of the experiencer (*mi sembra che*) yields a different semantics (see Section 2.1), we will also take into account the presence of the experiencer to assess if *mi sembra che* shows some construction-specific constraints or preferences in relation to its subjects or if it behaves similarly to *sembra che*.

3.2 Corpora

The analysis will be carried out on both written and spoken data, for two reasons. The first is to fill a gap in the study of these constructions in Italian, since previous research has focused only on written Italian (Kratschmer 2006; Musi 2015). The second reason is that, often, the study of “traditional” syntactic structures has been based on data coming from written language or from intuition. This approach leads to a biased perspective on grammar (Linell 1982; Hilpert 2019): if features typical of spoken language are not taken into account, the generalizations made on a phenomenon are at least incomplete.

The written data have been 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 150M words. We extracted a random sample of 622 occurrences of the verb *sembrare*⁵. Among these occurrences, only 216 turned out to be instances of *sembrare*+infinitive or *sembra che*.

As for spoken Italian, two corpora were used: LIP (De Mauro et al. 1993) and KIParla (Mauri et al. 2019). These two spoken corpora include data from different communicative situations, ranging from free dialogues to radio and TV speech. At the time when the research was carried out, the two corpora put together contained 1.5M words. The total occurrences of *sembrare* were 950, but only 189 of them turned out to be instances of *sembrare*+infinitive or *sembra che*.

We excluded from the samples 3 occurrences from the written corpus and 8 occurrences from the spoken corpora. These occurrences had been identified instances of the *sembra che* type, but their subordinate clauses did not have a noun phrase as their subject: they were either impersonal (in (7) we find an impersonal *si*) or had a clausal subject (8).

- (7) *sembra [...] che si arrivi ad esempio al collegio con delle moszioni [...] smussate*
‘for example, it seems that we go to the board with blunted motions.’ (LIP)

- (8) *sembra che non sia affatto detto che il copista [...] sia quel giovane*
‘it doesn’t seem certain that young man is the copyist.’ (KIParla)

Thus, the final dataset for the analysis consists of 394 total occurrences: 213 from the written corpus, and 181 from the spoken corpora.

⁵ We extracted a sample of 622 occurrences from CORIS to match the number of occurrences of *sembrare* contained in the KIParla corpus of spoken Italian. However, given the discrepancy in the number of constructions coded as *sembra che* and *sembrare*+infinitive in the written and spoken dataset (216 vs. 110), data from LIP corpus of spoken Italian were added to the KIParla ones.

3.3 Annotation

All examples in the dataset were annotated for the following variables: subject length (Section 3.3.1), animacy (Section 3.3.2) and referent accessibility and type of anaphor (Section 3.3.3).

3.3.1 Subject length

Noun phrase length has been proven to influence speakers' choice between syntactic alternants both in formal (Ross 1967) and in functional studies (Thompson 1990; Hawkins 1994). In particular, it has been claimed to affect the linear order of the constituents: typically, longer and more complex elements tend to be placed towards the end of the clause, and tend to follow shorter constituents. For instance, Construction Grammar accounts of the verb-particle alternation in English showed that when the direct object NP is longer, the continuous construction (9a) is preferred over the discontinuous one (9b) (Gries 1999; 2003).

- (9) a. *He **brought back** the books that he had left at home for so long.*
b. *??He **brought** the books that he had left at home for so long **back**.*
(Gries 1999:110)

As we saw in Section 2.2, Ruppenhofer (2004) found that in spoken English “raised” subjects of the verb *seem* are shorter than the unraised ones. We hypothesize that the same applies both to written and spoken Italian, as we suppose that the subject NP of *sembrare*+infinitive is more easily identifiable, and thus less formally marked.

We measured subject NP length in terms of number of words: this has been proven to be an effective and fairly simple measure of NPs' syntactic complexity in other studies (Wasow 1997; Szmrecsányi 2004).

3.3.2 Animacy

Animacy is a widely studied semantic feature of noun phrases. It is known to affect a vast number of morphosyntactic phenomena in many languages, ranging from case marking to subject selection (for an overview, cfr. Yamamoto 1999: Chapter 2). As for syntactic alternations, Bresnan et al. (2007) found that in the dative alternation (mentioned in Section 3.1) an animate recipient favours the use of the ditransitive construction.

Some studies also claim that speakers are more likely to “promote” animate NPs to the first position in the clause (e.g., through the use of passive constructions, topicalizations etc.). However, research on the matter provides mixed results (Ransom 1977; Harris 1978; Snider & Zaenen 2006).

We coded all occurrences for animacy, distinguishing three values: Human, Animate, and Non-Animate. Our hypothesis is that *sembrare*+infinitive subjects tend to be higher on the animacy scale (i.e., Human, Animate) than *sembra che* ones.

3.3.3 Referent accessibility and type of anaphor

As we saw in Section 2.2, information status has been used to account for the position of raised subjects. More generally, it constitutes a relevant discourse-functional correlate in language production, as research on syntactic alternations has shown (e.g. Gries 2003; Bresnan et al. 2007).

The vast literature on information status highlighted that the distinction between given and new (or active and inactive) referents is not sufficient, so the notion of inferrability was introduced (Prince 1981; Chafe 1987; Lambrecht 1994; Gries 1999). A referent can be inferred even if it has not been mentioned in the previous discourse: for instance, it can be evoked by some other referent in the linguistic or extra-linguistic context (Prince 1981; Lambrecht 1994: 113). In general, it has been noted that inferrable referents often behave similarly to referents already evoked in discourse (e.g. Birner 1994), even if there is not a complete overlap between the two (Prince 1992).

By virtue of these observations, we chose to follow the annotation scheme by Nissim et al. (2004) to code referent accessibility of the subjects in our occurrences (Referent accessibility). This annotation scheme includes three values, namely: Old, Mediated (i.e., inferrable), and New.

Another advantage of this scheme is that the Old and Mediated values include various subtypes of anaphoric links between the subject and its antecedent (for the full scheme, see the Appendix), which we took into account in the annotation (Type of anaphor).

3.4 Statistical methods

Since we tested the influence of the annotated variables (illustrated in Section 3.3) on the choice of the *sembrare* constructions, these variables were treated as independent variables in the statistical tests we used.

First, we tested the choice between *sembra che* and *sembrare*+infinitive, and then we divided *sembra che* into two subgroups (*sembra che* and *mi sembra che*) to assess whether there are differences between them or if they share the same constraints. We applied this procedure both to written and to spoken data separately.

Subject length was treated as an ordinal independent variable since word length generally does not follow a normal distribution, and was tested using the Mann-Whitney test (Ruppenhofer 2004: 237; Stefanowitsch 2020: 338). As for the categorical independent variables, i.e. Animacy, Referent accessibility and Type of anaphor, we tested their influence on the choice of the constructions using chi-squared tests. Since in some contingency tables more than 20% of the cells showed a value of less than 5, we simulated the p-value through 10.000 iterations of the Monte Carlo method (Desagulier 2017: 186). As post-hoc analysis, we looked at the standardized residuals of the cells: cells in the contingency table that showed a value exceeding ± 2 were assumed to contribute to the significance of the test (Agresti 2007).

4. Results

This section shows the results of our analysis for each coded variable. However, for the sake of brevity, we offer an accurate description of the data only in the case of significant

results, while we report the corresponding p-value for all the other tests.

4.1 Distribution of the constructions

Before delving into the features of the subjects, we looked at the distribution of the constructions in written and spoken language. As we see in Table 1, *sembrare*+infinitive is more used than *sembra che* in written Italian, while it is very rarely used in spoken Italian.

	<i>sembra che</i>	<i>sembrare</i> +infinitive	Total
Written	85 (39,4%)	131 (60,6%)	216 (100%)
Spoken	163 (86,2%)	26 (13,8%)	189 (100%)
Total	248	157	405

Table 1. Frequency of *sembra che* and *sembrare*+infinitive.

It could seem from this distributions that *sembra che* is particularly favoured in spoken contexts. However, if we separate the occurrences with overt experiencer, we notice that it is the *mi sembra che* construction that shows an increase in the spoken data (100 vs. 24): it constitutes the majority of the occurrences of the “impersonal construction”, while *sembra che* without experiencer shows a similar distribution between written and spoken language (Table 2).

	<i>sembra che</i>			<i>sembrare</i> +infinitive		
	<i>mi sembra che</i>	<i>sembra che</i>	Total	<i>mi sembra+inf</i>	<i>sembrare+inf</i>	Total
Written	24	61	85	1	130	131
Spoken	100	63	163	1	25	26
Total	124	124	248	2	155	157

Table 2. Frequency of the constructions with and without the experiencer.

Furthermore, it is noteworthy that *sembrare*+infinitive almost never appears with the experiencer in our corpus (*mi sembra*+infinitive shows up only twice). This behaviour is very similar to the one displayed by *parecer*+infinitive in Spanish (Davies 1997: 55-56; Cornillie 2007: 80-82). Since we have only two occurrences of this construction, we did not test them separately from *sembrare*+infinitive in our analysis.

4.2 Subject length

The first feature we investigated is the length of the subject NPs.

In the data from the written corpus, we notice a difference in the subject lengths of the two constructions. Specifically, they diverge with respect to their mean ranks, that result to be higher for *sembrare*+infinitive (*sembra che* = 2.41; *sembrare*+infinitive = 3.04): in other words, the subjects used with *sembrare*+infinitive tend to be longer than the ones used with *sembra che*. If we compare the distributions of the subject lengths of the two constructions, we can see that the greatest difference is to be attributed to the group of

zero subjects (Figure 2): while zero subjects are the most numerous group among *sembra che* subjects, they are less used than one and two-word subjects with *sembrare*+infinitive. This difference resulted to be highly significant (p-value = 0.005).

We notice the same distribution also in the spoken data, even though the subjects tend to be generally shorter (Figure 3). The result of the test was significant, although the p-value was lower than the one from the test on the written data (p-value = 0.025). Indeed, in this case, the mean ranks of the two distributions are closer than in the written data (*sembra che* = 1.58; *sembrare*+infinitive = 1.64), and this could be the reason for this result.

Regarding the comparison between *sembra che* and *mi sembra che*, the mean ranks of the subject lengths of the two constructions turned out to be very close, both in the written (*sembra che* = 2.32; *mi sembra che* = 2.65) and in the spoken data (*sembra che* = 1.52; *mi sembra che* = 1.51). Coherently with this tendency and with the similarity of the shape of the distribution curves, the tests did not yield significant results (p-value for written data = 0.22; p-value for spoken data = 0.23).

4.3 Animacy

The distribution of the animacy values in the corpora shows that *sembrare*+infinitive co-occurs mainly with Non-Animate subjects, both in written and spoken data. The same holds for *sembra che* in the written corpus, while in the spoken corpus this construction appears mostly with Human subjects.

However, the animacy of the subjects turned out to be not significant both in written (p-value = 0.061) and in spoken data (p-value = 0.068).

We also looked for a possible difference between *sembra che* and *mi sembra che*, but even this test did not yield significant results (p-value for written data = 0.48, p-value for spoken data = 0.24).

4.4 Referent accessibility

As for Referent accessibility, most of the NPs turned out to be Old, both in written and spoken data. Nonetheless, we decided to test the distributions to assess if subtle, but significant differences could be found.

Our results did not show any substantial difference between *sembrare*+infinitive and *sembra che* in the accessibility of their subjects (p-value for written data = 0.56, p-value for spoken data = 0.18), even if *sembrare*+infinitive does not select any New subject in the spoken data.

Similar results were found in testing *sembra che* and *mi sembra che*. Although in the written corpus *mi sembra che* showed a higher number of New subjects as compared to *sembra che*, this association didn't turn out to be significant (p-value = 0.063). As for the spoken corpus, we did not find any relevant tendency in our data (p-value = 0.39).

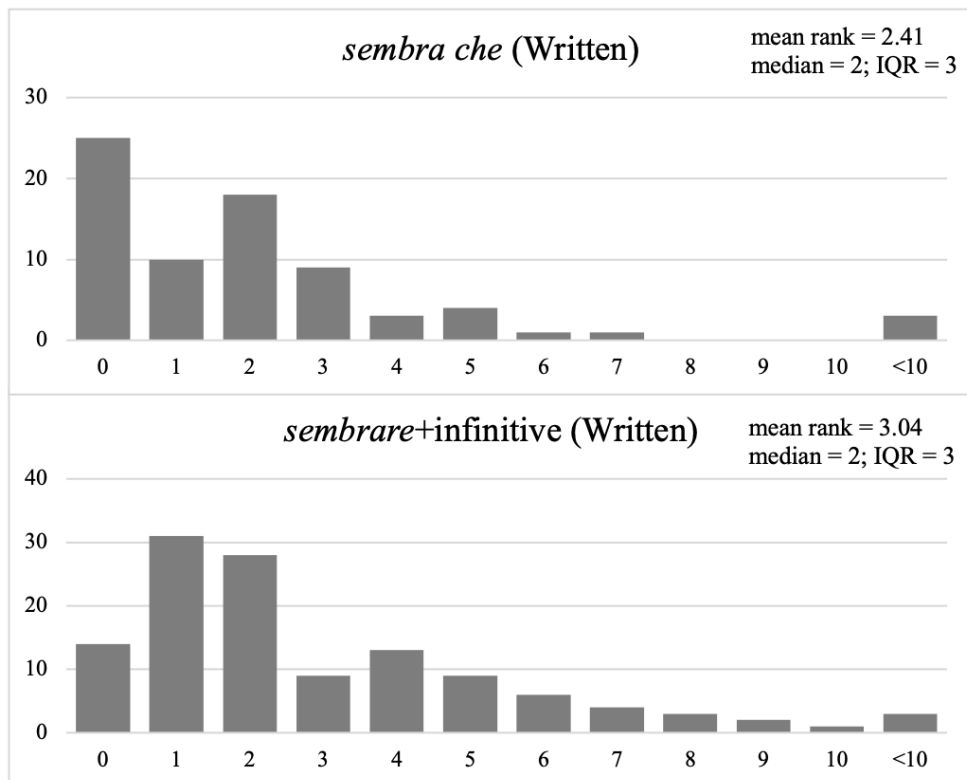


Figure 2. Subject length of *sembra che* and *sembrare+infinitive* in the written corpus.

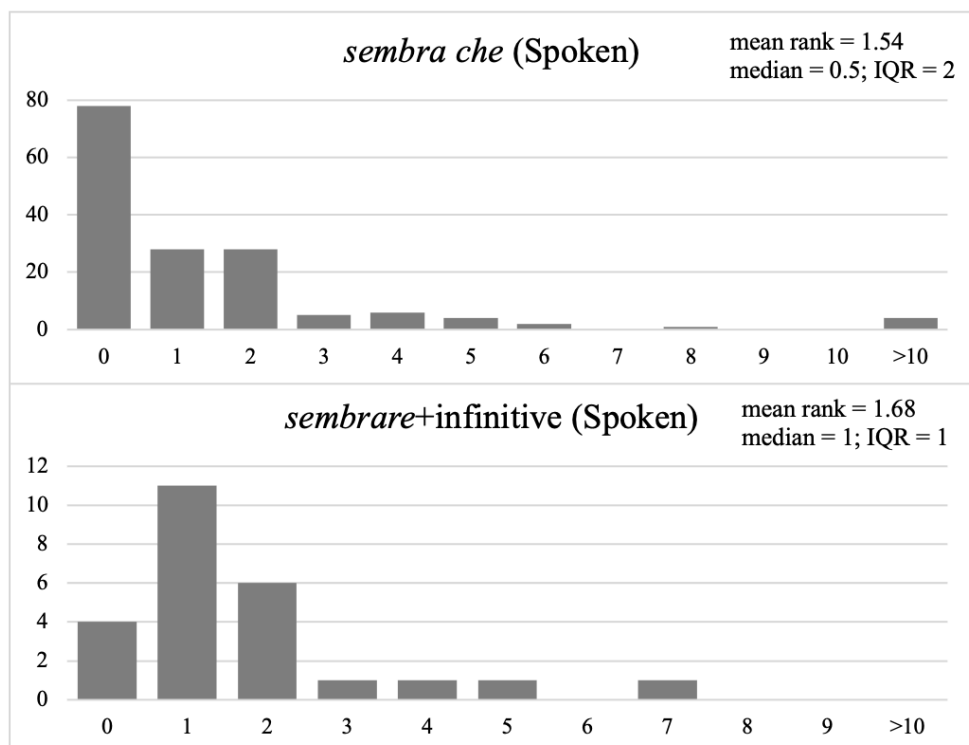


Figure 3. Subject length of *sembra che* and *sembrare+infinitive* in the spoken corpus.

4.5 Type of anaphor

Finally, we turn to the analysis of the type of anaphoric link between the subjects and their antecedents.

In the written corpus, two values (Old_general and Old_generic) stand in contrast with the overall frequency of the two constructions: both occur almost only with *sembra che*. However, we found no significant effect of the type of anaphoric link on the choice of constructions (p-value = 0.12), probably due to the low number of occurrences of these types of subjects in the corpus. Similarly, the comparison between *sembra che* and *mi sembra che* in the written corpus did not return significant results (p-value = 0.25).

On the contrary, if we look at the data from the spoken corpus in Table 3, some differences in the distributions emerge. For instance, the most evident one is that even though *sembrare*+infinitive is less frequent than *sembra che*, it has a higher number of relative pronouns (Old_rel) as subjects.

The results of the statistical test confirmed a highly significant association between the type of anaphoric link and the choice of the construction (p-value = 9.999e-05). The analysis of the standardized residuals showed that the two levels that contribute mostly to the significance of the test are Old_general (1st and 2nd person pronouns) and Old_rel, respectively associated with *sembra che* and *sembrare*+infinitive (Figure 4).

Moreover, we found a difference in the selection of the subject between *sembra che* and *mi sembra che*. As we can see in Table 3, Old_generic subjects (i.e., generic pronouns and NPs) are used only with *sembra che*. This discrepancy determined the significance of the test (p-value = 0.003), as the analysis of the residuals testifies (Figure 5).

		<i>sembra che</i>			<i>sembrare</i> +infinitive	Total
		<i>sembra che</i>	<i>mi sembra che</i>	Total		
Old	event	2	7	9	0	141
	general	7	15	22	0	
	generic	8	0	8	1	
	ident	28	49	77	13	
	rel	0	2	2	9	
	Total	44	74	118	23	
Mediated	event	1	0	1	0	21
	general	1	6	7	0	
	poss	0	1	1	0	
	set	1	1	2	0	
	sit	1	7	8	2	
	Total	4	15	19	2	
Total		137			25	162

Table 3. Types of anaphoric links in the spoken corpus.

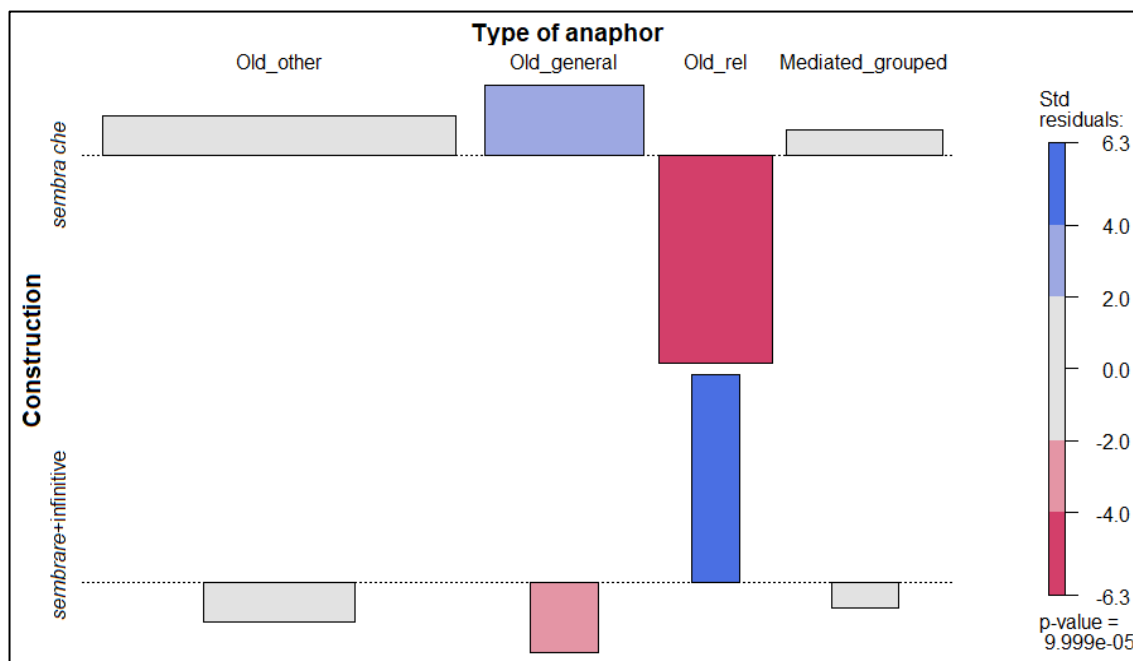


Figure 4. Association of *sembra che* and *sembrare*+infinitive with the type of anaphor.⁶

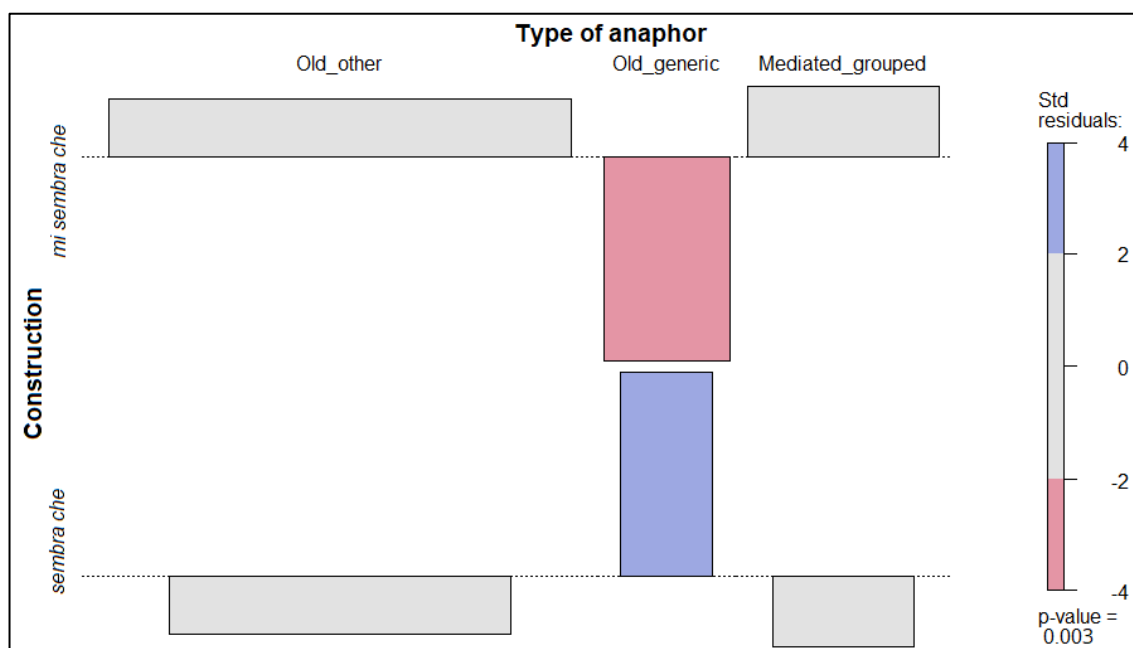


Figure 5. Association of *sembra che* and *mi sembra che* with the type of anaphor.

⁶ We grouped the levels that do not contribute to the significance of the test. The association plot was realized using a modified version of the *assocplot()* function of the R package *vcd* (Meyer et al. 2006). The modified function is available at: https://github.com/fla-pi/associationplot_stdresiduals

5. Discussion

5.1 Subject length: a hint at topicality?

The first difference we found between *sembra che* and *sembrare*+infinitive is that *sembrare*+infinitive is significantly less used with zero subjects (i.e., when the subject is omitted). This diverges from previous studies on these constructions: in her research, Musi shows that subjects are omitted only with *sembra che* (2015: 275-276). However, this discrepancy could be explained by differences in the composition and in the dimension of the corpora: Musi (2015) focuses only on data coming from reviews and opinion articles and takes into account a lower number of occurrences as compared to our study.

A possible way to account for this result has to do with the relationship between word length and topicality. However, we first need to give an operational definition of topicality. According to Givón, a topical referent is characterized by its recurrence in a larger portion of the discourse. Referent recurrence has two sides: an anaphoric one (its presence in the previous discourse, i.e., its accessibility), and a cataphoric one (its relevance to the subsequent discourse) (2001: 254).

If we focus on accessibility (the anaphoric aspect of topicality), we expect the choice of a shorter NP (or zero NP, as in our case) to correlate with the easier retrievability of its referent, as also highlighted by Givón's topic continuity scale (1983: 359):

- (10) *zero > clitic pronoun > independent pronoun > full DEF-NP > modified DEF-NP*

Following this scale, an increase in formal markedness should correlate with a decrease in the predictability (i.e., accessibility) of a referent. In our case, the prediction would be that the subjects of *sembra che* are more accessible than the subjects of *sembrare*+infinitive. Our analysis, however, did not find Referent accessibility to be a distinguishing factor between the two constructions: indeed, the majority of the subjects of *sembrare*+infinitive turned out to be discourse-old even if they are generally more formally marked than *sembra che* ones. Thus, at least in our case, it would be incorrect to resort to the anaphoric side of topicality to explain the choice between the two constructions and the difference in the length of their subjects.

Instead, the tendency of *sembrare*+infinitive to attract more formally marked subjects could be linked to the cataphoric side of topicality. Going back to the scale in (10), it has been noted that more marked forms generally indicate an element as more prominent (Baloghne Nagy 2013: 20). Prominence can be linked to the cataphoric aspect of topicality: for instance, prominent elements are more likely to be mentioned in the following discourse (von Heusinger & Schumacher 2019). In our corpus, we find instances of *sembrare*+infinitive in which the subject of the construction (a modified definite NP) is referred to several times in the following discourse:

- (11) *questo corpo nudo sembra reagire agli avvenimenti del mondo. Le sue malattie infantili [...] sono come reazioni cutanee alla follia universale, e la sua asma, poi, tutte quelle manifestazioni allergiche, quella sua pelle così vulcanica...È impossibile dissociare le peripezie della Storia dai tormenti di quel corpo [...]*
'This naked body seems to react to the events of the world. Its childhood diseases are like skin reactions to universal madness, and its asthma, all those allergic

manifestations, **its** volcanic skin...It is impossible to separate the vicissitudes of History from the torments of **that body**.' (CORIS)

Thus, we could hypothesize that the reason why the subjects of *sembrare*+infinitive tend to be overtly expressed more often than the *sembra che* ones is that they are more prominent, or topical in a cataphoric sense. In this way, *sembrare*+infinitive can be considered similar to information structure constructions for its topicalizing function (as proposed in Givón 2001: 276).

Future studies may test this hypothesis by assessing various aspects, such as the persistence of the referent expressed by the subject NP. If confirmed, it would also help to explain the more frequent use of *sembrare*+infinitive in written language: in absence of other means (e.g., prosody), promoting an NP to a fronted position is necessary to mark its topicality (Gundel 1985: 95).

5.2 Relative pronouns: antecedent distance and ease of processing

In the previous section, we proposed a topicalizing function for *sembrare*+infinitive, as suggested by the use of longer, more marked subjects. However, our results show that *sembrare*+infinitive is favoured also by a type of less prominent subjects, namely relative pronouns (12).

- (12) [...] *a un estremo c'è la musica che sembra avere solo un valore estetico* [...] 'at one end there is music that seems to have only an aesthetic value' (KIParla)

In contrast, *sembra che* is almost never chosen when the subject is a relative pronoun, excluding very rare exceptions:

- (13) [...] *c'è un bimbo [...] che sembra che stia giocando* 'there is a baby that seems to be (lit. 'that is') playing' (KIParla)

This tendency was found to be significant in the spoken data, but a closer inspection shows that it is also valid for the written data; perhaps the non-significance of the test conducted on the latter is due to the fact that *sembrare*+infinitive is very frequent. In the spoken corpus, on the other hand, relative clauses are an environment that particularly favours the rare use of *sembrare*+infinitive. Interestingly, the preference for the infinitival constructions in relative contexts was noticed also for Spanish (Holmlander 2004) and English (Lenardič & Ilc 2019).

According to Kratschmer (2006), the use of *sembrare*+infinitive in a relative clause would have the function of thematizing an element introduced as a rheme in the main clause, as we see in (12). However, as Holmlander (2004: 18) notes for Spanish, it is more plausible that in these cases *sembrare*+infinitive does not have a topic-marking function, because the subject is already structurally in initial position. Moreover, the local topic is already marked by the antecedent of the relative pronoun (Langacker 1993: 34; Van Valin 1995: 52). Thus, we assume that in this case *sembrare*+infinitive is not used to mark the subject as a topical element. On the contrary, the use of a relative pronoun expresses a strong topic continuity with the referent just evoked in the main clause.

Rather, this preference can be attributed to the position of the subject of *sembrare*+infinitive. It is known in the literature that the relativization of a given element (e.g., the subject) is easier when it belongs to a main clause rather than a subordinate clause (Comrie 1989: 161). The reason probably pertains to the processing cost: a construction like *sembrare*+infinitive embedded in a subject relative clause is easier to process than the corresponding construction with *sembra che*, because it contains less lexical material intervening between the noun modified by the relative clause (the filler) and its position within the relative clause (the gap) (Gibson 1998). In the case of *sembra che*, the effort and the memory load required to relate the filler to the gap in the relative clause would be greater:

- (14) a. *c'è la musica* [filler] *che* ___ [gap] *sembra avere solo un valore estetico*
 ‘there is music [filler] that ___ [gap] seems to have only an aesthetic value.’
 b. *c'è un bimbo* [filler] *che sembra che* ___ [gap] *stia giocando*
 lit: ‘there is a baby [filler] that seems that ___ [gap] is playing.’

Thus, the function of *sembrare*+infinitive in these occurrences is to promote the subject to the main clause, in order to make it more accessible for relativization.

5.3 Marking the interlocutor(s): the hic et nunc of speech

Another trend concerns 1st and 2nd person subjects, that is, all those referring to the speaker and interlocutor(s). Unlike *sembra che* (cfr. (14)-(15)), we did not find occurrences of *sembrare*+infinitive with non-3rd person subjects (singular or plural). This phenomenon is common both to the written and the spoken corpora but is more prominent in the latter. The reason is straightforward: the reference to the interlocutors is more common in speech situations (Chafe 1985).

- (14) *cioè sembra che mi stai intervistan[do]*
 ‘I mean, it seems you are interviewing me.’ (KIParla)

 (15) *sembra che ora ti devo rivelare chissà che cosa*
 ‘it seems that I’m going to reveal you who knows what’ (KIParla)

This constraint on infinitive constructions with SEEM verbs has also been noted in some studies on Spanish (Bolinger 1961; 1972; Davies 1997), whereas in English, for example, it is common to find 1st and 2nd person subjects with *seem*+infinitive (Davies 1997: 44):

- (16) *I seem to have offended Mary.*

This behaviour is not unexpected in the Italian epistemic system. Pietrandrea (2005: 122-123) shows that the epistemic reading of Italian modals and of the epistemic future is not possible, or it is at least awkward, when the subject coincides with the speaker or the interlocutor(s). If the speaker is the one who evaluates the probability of a state of affairs, they should have high epistemic control over their (or their interlocutors’) actions or states, thus making it “useless to place these actions and states within an epistemic domain made up of beliefs, rather than facts” (Pietrandrea 2005: 120). Therefore, when the

speaker has low epistemic control over the state expressed by the verb in the infinitive (17) this constraint is violable.

- (17) *Devo essere scema.*
'I must be an idiot.' (Pietrandrea 2005: 120)

Thus, our results suggest a difference between *sembrare*+infinitive and *sembra che* with respect to the *origo* of the evaluation, i.e., the experiencer from whose perspective the proposition is evaluated. The use of *sembrare*+infinitive implies that the unexpressed experiencer coincides with the speaker (and this might explain why the experiencer is not overtly marked with *sembrare*+infinitive, as we saw in Section 4.1).

Instead, the use of *sembra che* makes it possible to interpret the unexpressed experiencer as a third, often arbitrary, person. Such an interpretation would permit the use of 1st and 2nd person subjects in the complement clause. This can be noted in (18), in which the addition of *ma non fumo* 'but I don't smoke' makes the state of affairs conveyed by the complement clause of *sembra che* counterfactual, attributing the (wrong) evaluation to an arbitrary observer:

- (18) [...] *io sembra che fumo ma non fumo*
'me, it seems that I smoke, but I don't.' (LIP)

This explanation is confirmed by the (rare) occurrences of *sembrare*+infinitive with 1st person subjects found in CORIS. In example (19), it is plausible that the speaker does not believe to have epistemic control over their mental state:

- (19) *So che è irrazionale, ma quando si tratta di lei, sembro perdere ogni senso della misura.*
'I know it is irrational, but when it comes to her, I seem to lose all sense of proportion.' (CORIS)

However, the coincidence of the experiencer and the subject of the proposition is not enough to explain this distribution. As a matter of fact, *mi sembra che* (whose experiencer is unequivocally the speaker) can be used without any problem with 1st person subjects:

- (20) [...] *a me mi sembra che devo ragionarci prima di fare una scelta*
'It seems to me that I have to think about it before making a choice.' (KIParla)
- (21) *mi sembra che io [...] su tante questioni di merito [...] non posso entrarci*
'It seems to me that, as for many things, I can't get into matters of substance.'
(LIP)

The explanation lies in another element, namely the coincidence of the moments of knowledge acquisition and enunciation. As we saw in Section 2.1, *sembrare*+infinitive is mainly used to express inferences based on the speaker's reasoning. It has been claimed that the speaker's inferences are generally drawn from events taking place at the moment of enunciation (Cornillie 2007: 28). Thus, in the case of 1st or 2nd person subjects, an inference would be drawn about some state of affairs regarding the participants to the

current speech situation, which would be awkward, as we saw, due to the strong epistemic control of the speaker over the situation.

In contrast, the occurrences of *mi sembra che* containing 1st person subjects often express judgments based on the speaker's belief (21). Even when the evaluation is based on an inference, it refers to a habitual state of affairs (20). Admittedly, it is more natural to express an uncertain evaluation based on a general belief or about a habitual state, than about an occasional event experienced by the speaker/interlocutor in a given speech situation. Interestingly, in (19) the acceptability of *sembrare*+infinitive could be attributed to the same reason: in this case too, the speaker is evaluating a habitual state of affairs.

In conclusion, the constraint shown by *sembrare*+infinitive on 1st and 2nd person subjects is probably due to its more frequent performative use (cfr. Nuyts 2001: 40), i.e., *sembrare*+infinitive is more often implicitly anchored to the participants and to the time of the speech situation than *(mi) sembra che*.

5.4 Generic 'you' and experiencer defocusing

The last result we analyse is not related to *sembrare*+infinitive, but it highlights a difference between *sembra che* and *mi sembra che*. In the spoken data, it emerged that generic subjects are used only with *sembra che*. Among these subjects, we found either generic NPs (*la gente* 'the people') (22), or indefinite pronouns (23), or 2nd person singular subjects (generic 'you') (24).

- (22) [...] *sembra che la gente ci va con le valigie di libri*
'It seems that people go there with bags full of books.' (KIParla)
- (23) TOI019: *sembra che non ti guardava nessuno e poi ti guardavano*
'It seems like no one was looking at you, and instead, they were looking at you'
TOR004: *eh ma in paese è sempre un po' così*
'Well, it's always a little bit like that in villages' (KIParla)
- (24) *lo so partir da sola [...] sembra tipo che non hai nessuno a cui chiedere*
'I know, leaving alone [...] it seems like you have no one to ask.' (KIParla)

While in (22) the use of *mi sembra che* would be acceptable, in the other examples the insertion of the experiencer seems intuitively less natural. On a closer inspection of the occurrences, we hypothesize that this may be an effect of the conceptual space associated with the use of generic 'you'. As a matter of fact, in most of our occurrences (5 out of 7) the generic 2nd person pronoun is present in the utterance, even when it does not have a subject function (23).

If we compare (22) and (24), they both contain strategies of agent defocusing⁷ (Sansò 2006) but type of defocusing is different. In (22) the agent is underspecified, but the event encoded takes place factually; on the contrary, in (24), the agent is "virtually all humanity" and the event encoded "does not correspond directly with a contingent real event" (Sansò 2006: 264). More specifically, the function assumed by generic 'you' in

⁷ "Agent" is to be intended here as a macro-role (Sansò 2006: 237).

our occurrences is to generalize the speaker's experience and point of view by attributing them to anyone who might virtually be in a similar situation. In order to do so, the speaker hides their personal point of view behind a non-referential 2nd person pronoun (Kluge 2016; Auer & Stukenbrock 2018), setting up a fictive conceptualizer in the mental space of the addressee (Rubba 1996).

We suppose that the introduction of a deictic element referring to the speaker (the experiencer) could clash with this attempt to generalize the judgment, reconnecting it to an actual speaker. On the contrary, *sembra che* defocuses the experiencer, allowing for its arbitrary reading. This probably constructs the judgment as universal and unties it from the speaker's personal beliefs, marking it as more intersubjective. In (23) and (24) the omitted experiencer could be a generic *ti* 'to you'. A question to be addressed with more data is why we did not find any generic 'you' as experiencer in these occurrences and whether its usage would be as felicitous as the use of *sembra che* without experiencer.

In conclusion, we hypothesize that the presence in the context of a generic 'you' disfavours the use of *mi sembra che*. We wonder if the same holds for *sembrare*+infinitive, whose use generally implies the presence of the speaker as evaluator (as we saw in Section 5.3), but the relative rarity of these contexts does not allow us to draw any conclusions, and further studies on the matter are needed.

5.5 Summing up: a constructionist perspective

In our analysis on *sembrare*+infinitive and *sembra che*, we found various constraints and preferences on subject selection. We summarize them in Table 4.

<i>sembrare</i> +infinitive	<i>sembra che</i>
<ul style="list-style-type: none"> - Preference for longer subjects (tendency not to omit the subject). - The referent of the subject cannot correspond to the speaker or to the addressee(s). 	<ul style="list-style-type: none"> - The subject (generally) cannot be a relative pronoun.

Table 4. Constraints and preferences encoded in *sembrare*+infinitive and *sembra che*.

One of these constraints could be accounted for by the traditional formal analyses of this alternation: indeed, the preference for *sembrare*+infinitive in relative clauses has been accounted for from a generativist point of view (Lenardič & Ilc 2019).

Nevertheless, a formalist view on this alternation would hardly be able to explain a constraint such as the one on 1st and 2nd person subjects. In fact, this constraint is the one that most undermines the assumed "transparency" of *sembrare* (see Section 1): if some subjects are not good candidates to be "raised", it means that also *sembrare* plays some role in the selection of the subject.

Instead, Construction Grammar offers a model more suited to represent and explain these facts: it makes it possible to specify, within the constructions, the functional information needed to account for these constraints. As a matter of fact, based on the

constraints we found, we identified some distinctive functional aspects of these constructions that could motivate them⁸ (Table 5).

	<i>sembrare</i> +infinitive	<i>sembra che</i>
FORM	N _x <i>sembrare</i> V _y	<i>sembrare che</i> N _x V _y
FUNCTION	evaluation, based on indirect evidence, that X is in the situation described by Y	
Information structure		
- (Possibly) Topic-marking function: X is relevant for the subsequent discourse.	-	
Degree of performativity		
- The evaluation is (by default) made by the speaker and is based on some event simultaneous to the communicative situation.	- The evaluation is not necessarily made or endorsed by the speaker. - The evaluation can be referred to fictive events and can be attributed to a fictive experiencer.	

Table 5. Functional information encoded in *sembrare*+infinitive and *sembra che*.

However, we should not forget that *sembrare*+infinitive and *sembra che* share many aspects of their semantics, to the point that they can be used interchangeably in various contexts (Section 1). We claim that the concept of allostruction, introduced in Section 3.1, makes it possible to account for these data. From the formal point of view, *sembrare*+infinitive and *sembra che* are quite different constructions, but still, they share a semantic space, namely their general evidential type (Table 5). This fact could be captured by postulating that the common aspects are encoded in a constructeme (speaker’s knowledge of the quasi-synonymy of the two forms). At the same time, however, they are not entirely interchangeable because, from a pragmatic point of view, they serve different functions – concerning information structure and the degree of performativity – and got specialized for different niches (e.g., relative clauses, 1st person subjects etc.).

A final question concerns the status of *mi sembra che*: should we consider it an allostructional variant of *sembra che* and *sembrare*+infinitive? In our view, there is no ground to postulate this relationship for at least two reasons.

The first is that given the formal similarity between *sembra che* and *mi sembra che*, it would be useless to postulate a formally underspecified constructeme. Instead, it is more

⁸ We did not include functional features to explain the relative pronoun constraint, because we hypothesized the influence of processing factors (without postulating information structural or pragmatic motivations). Nonetheless, we could add to *sembrare*+infinitive the function of making the subject more accessible for relativization.

likely that *mi sembra che* results from the unification (Kay 2002) of *sembra che* and the dative experiential construction *mi/a me* ‘to me’.

The second reason is that, according to the literature (Section 2.1) and looking at the occurrences analyzed, *mi sembra che* has a slightly different semantics from those of *sembra che* and *sembrare*+infinitive: often, rather than highlighting the evidential semantics, it mainly marks the epistemic evaluation made by the speaker, putting aside the sources based on which the evaluation is made. Therefore, under the current interpretation, it is plausible that the unification with the dative experiencer construction *mi* partially overrides the evidential meaning, privileging the uncertainty on the likelihood of the presented state of affairs.

This analysis allows us to explain the behaviour displayed by *mi sembra che* without recurring to more abstract objects. We propose that:

- Most of the preferences displayed by *mi sembra che* are inherited from *sembra che* (subject length, relative clause contexts, 1st and 2nd person subjects).
- The unification with the dative experiencer construction *mi* adds the constraint on fictive contexts introduced through the use of generic ‘you’.

However, further studies, both on the nature of horizontal links and on the function of *mi sembra che*, will further explore the validity of this interpretation.

6. Conclusion

Our study showed that the alternation between *sembra che* and *sembrare*+infinitive is influenced by various features of the NPs that occupy the subject slots of the constructions. This result supports the idea that it is not sufficient to describe these two constructions as mere paraphrases of each other and that they are not totally used in free variation. Rather, there are linguistic niches in which speakers prefer to use one of the two variants. This situation can be fruitfully described as an allostructional relationship in the framework of Construction Grammar: there is a substantial overlap between the semantics of *sembra che* and *sembrare*+infinitive to keep them together, but at the same time they differ with regard to some other functional aspects and their distribution.

Furthermore, we found an interesting contrast between *sembra che* and *mi sembra che*, since the latter is generally not used in fictive contexts. In this case, we chose not to analyze the relationship between the two constructions in allostructional terms, because we judged the unification of *mi* and *sembra che* to be enough to account for this distributional difference.

On a final note, we want to stress the centrality of spoken data in our study. If we had ignored spoken data, we would not have found most of the significant constraints in our analysis despite the fact that they could also be applied to written-language uses of *sembrare*-constructions. In our case, there are various reasons for these differences between written and spoken data: lower frequency of some phenomena in written language (reference to the interlocutors, use of the generic ‘you’), or less relevance of some constraints in determining the frequency of a construction in our written data (relative clauses). The relevance of the constraints found in our study, however, reveals that the traditional, intuition- or written-based approach to syntax runs the risk of incurring into incomplete or misleading generalizations, and that this risk can be avoided

only by taking into account both written and spoken data in the study of syntactic phenomena.

Appendix

The table illustrates scheme the criteria for the annotation of the type of anaphoric link (Nissim et al. 2004), as described in the annotation guidelines in Nissim (2003).

Referent Accessibility	Subtypes	Description
Old	identity	The anaphor and the antecedent are coreferential, i.e. they refer to the same entity
	relative	Relative pronouns
	generic	Any generic use of pronouns
	ident_generic	Generic pronouns that are coreferential with an already introduced generic pronoun
	general	“I” and “you” are used in their referential form
	event	NPs that refer to events
Mediated	bound	Bound non-coreferential pronouns
	general	The entity is mediated on general knowledge grounds
	event	The entity is associated with an event that has been previously mentioned
	aggregated	Coordinated NPs that involve at least one mediated and/or old entity
	func_value	The relationship between a function and its value
	set	Any kind of set relations
	poss	Possessive relation
	part	Part-whole relation for physical objects
	sit	The anaphor is part of a situation set up by the antecedent
New	-	-

Bibliography

- Achard, Michel. 2000. The distribution of French raising constructions. In Conathan, Lisa J. & Good, Jeff & Kavitskaya, Darya & Wulf, Alyssa B. & Yu, Alan C.L. (eds.), *Proceedings of the 26th Annual Meeting of the Berkeley Linguistics Society, Berkeley, 18-21 February 2000*, 1–15. Ann Arbor: Sheridan Books.
- Agresti, Alan. 2007. *An Introduction to Categorical Data Analysis*. 2nd ed. Hoboken: John Wiley & Sons.
- Aijmer, Karin. 2009. *Seem* and evidentiality. *Functions of Language* 16(1). 63–88.
- Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.
- Audring, Jenny. 2019. Mothers or sisters? The encoding of morphological knowledge. *Word Structure* 12(3). 274–296.

- Auer, Peter & Stukenbrock, Anja 2018. When ‘You’ Means ‘I’: The German 2nd Ps.Sg. Pronoun *du* between Genericity and Subjectivity. *Open Linguistics* 4(1). 280–309.
- Bach, Emmon 1977. Review of *On Raising: One Rule of English Grammar and Its Theoretical Implications*, by P. M. Postal. *Language* 53(3). 621–654.
- Baloghné Nagy, Gizella. 2013. *Left dislocation in Optimality Theory*. Budapest: Eötvös Loránd University. (PhD dissertation.)
- Barron, Julia. 2001. Perception and Raising Verbs: Synchronic and Diachronic Relationships. In Butt, Miriam & King, Tracy Holloway (eds.), *Time over Matter: diachronic perspectives on morphosyntax*, 73–104. Stanford: CSLI Publications.
- Birner, Betty J. 1994. Information status and Word Order: An Analysis of English Inversion. *Language* 70(2). 233–259.
- Bolinger, Dwight L. 1961. Syntactic Blends and Other Matters. *Language* 37(3). 366–381.
- Bolinger, Dwight L. 1968. Entailment and the Meaning of Structures. *Glossa* 2(2). 119–127.
- Bolinger, Dwight L. 1972. The syntax of *parecer*. In Valdman, Albert (ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*. 65–76. The Hague: Mouton.
- Bresnan, Joan & Cueni, Anna & Nikitina, Taytiana & Baayen, Harald R. 2007. Predicting the Dative Alternation. In Bouma, Gerlof & Kraemer, Irene & Zwarts, Joost (eds.), *Cognitive Foundations of Interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Arts and Sciences.
- Cappelle, Bert. 2006. Particle placement and the case for “allostructions”. *Constructions* 7. (<https://constructions.journals.hhu.de/article/view/381/396>) (Accessed on: 13.11.2023)
- Chafe, Wallace. 1985. Linguistic Differences Produced by Differences between Speaking and Writing. In Olson, David R. & Hildyard, Andrea & Torrance, Nancy (eds.), *Literacy, Language, and Learning*, 105–123. Cambridge: Cambridge University Press.
- Chafe, Wallace. 1987. Cognitive constraints on information flow. In Tomlin, Russell S. (ed.), *Coherence and Grounding in Discourse: Outcome of a Symposium, Eugene, Oregon, June 1984*, 21–51. Amsterdam-Philadelphia: John Benjamins.
- Comrie, Bernard. 1989. *Language Universals and Linguistic Typology: Syntax and Morphology*. 2nd ed. Oxford: Blackwell.
- Cornillie, Bert. 2007. *Evidentiality and Epistemic Modality in Spanish (Semi-) Auxiliaries: A Cognitive-Functional Approach*. Berlin: Mouton De Gruyter.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Davies, Mark. 1997. A Corpus-Based Analysis of Subject Raising in Modern Spanish. *Hispanic Linguistics*. 9(1). 33–63.
- Davies, William D. & Dubinsky, Stanley. 2004. *The grammar of raising and control: A course in syntactic argumentation*. Malden: Blackwell Publishing.
- De Mauro, Tullio & Mancini, Federico & Vedovelli, Massimo & Voghera, Miriam. 1993. *Lessico di frequenza dell’italiano parlato*. Milano: Etaslibri.
- De Vaere, Hilde & De Cuyper, Ludovic & Willems, Klaas. 2018. Alternating constructions with ditransitive *geben* in present day German. *Corpus Linguistics and Linguistic Theory*. 17(1). 73–107.
- Desagulier, Guillaume. 2017. *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics*. Cham: Springer.

- Dowty, David R. 1978. Governed transformations as lexical rules in a Montague Grammar. *Linguistic Inquiry* 9(3). 393–426.
- Fillmore, Charles J. & Kay, Paul & O'Connor, Mary C. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language* 64(3). 501–538.
- Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition*. 68. 1–76.
- Gisborne, Nikolas & Holmes, Jasper. 2007. A History of English Evidential Verbs of Appearance. *English Language and Linguistics* 11(1). 1–29.
- Givón, Talmy. 1983. Topic continuity in spoken English: An Introduction. In Talmy Givón (ed.), *Topic Continuity in Discourse: A quantitative cross-language study*, 343–364. Amsterdam-Philadelphia: John Benjamins.
- Givón, Talmy. 2001. *Syntax: An Introduction*, vol. II. Amsterdam-Philadelphia: John Benjamins.
- Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Gries, Stefan Th. 1999. Particle movement: A cognitive and functional approach. *Cognitive Linguistics* 10(2). 105–145.
- Gries, Stefan Th. 2003. *Multifactorial analysis in corpus linguistics: A study of particle placement*. London: Continuum.
- Grimm, Scott. 2010. An Empirical View on Raising to Subject. *ZAS Papers in Linguistics* 52. 83–109.
- Gundel, Jeanette. 1985. 'Shared knowledge' and topicality. *Journal of Pragmatics* 9(1). 83–107.
- Harris, Margaret. 1978. Noun animacy and the passive voice: A developmental approach. *The Quarterly Journal of Experimental Psychology* 30(3). 495–504.
- Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hilpert, Martin. 2019. *Construction Grammar and its application to English*. 2nd ed. Edinburgh: Edinburgh University Press.
- Hoffmann, Thomas & Trousdale, Graeme (eds.). 2013. *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press.
- Holmlander, Disa. 2004. Constructions with the raising verb *parecer* in Spanish. (Manuscript, Lund University.)
- Kay, Paul. 2002. An Informal Sketch of a Formal Architecture for Construction Grammar. *Grammars* 5. 1–19.
- Kratschmer, Alexandra R. 2006. Che te ne sembra? Semantica e pragmatica delle costruzioni italiane con sembrare/parere. In Olsen, Michel & Swiatek, Erik H. (eds.), *XVI Congreso de Romanistas Escandinavos/XVIe Congrès des Romanistes Scandinaves/XVI Congresso dei Romanisti Scandinavi/XVI Congresso dos Romanistas Escandinavos*. Roskilde: Roskilde Universitetscenter, Inst. f. Sprog og Kultur.
- Kluge, Bettina. 2016. Generic uses of the second person singular – how speakers deal with referential ambiguity and misunderstandings. *Pragmatics* 26(3). 501–522.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago: University of Chicago Press.

- Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press.
- Langacker, Ronald W. 1993. Reference-Point Constructions. *Cognitive Linguistics* 4(1). 1–38.
- Langacker, Ronald W. 1995. Raising and Transparency. *Language* 71(1). 1–62.
- Laporte, Samantha & Larsson, Tove & Goulart, Larissa. 2021. Testing the Principle of No Synonymy across levels of abstraction: A constructional account of subject extraposition. *Constructions and Frames* 13(2). 230–262.
- Leclercq, Benoît & Morin, Cameron. 2023. No equivalence: A new principle of no synonymy. *Constructions* 15(1). (<https://constructions.journals.hhu.de/article/view/535/584>) (Accessed on: 13.11.2023)
- Lenardič, Jacob & Ilc, Gašper. 2019. English raising predicates and (non-)finite clauses: diachronic and synchronic perspectives. *FLUMINENSIA* 31(1). 31–59.
- Linell, Per. 1982. *The Written Language Bias in Linguistics*. Linköping: University of Linköping.
- Mack, Jennifer E. 2010. *Information Structure and the Licensing of English Subjects*. New Haven: Yale University. (PhD dissertation.)
- Mauri, Caterina & Ballarè, Silvia & Gorla, Eugenio & Cerruti, Massimo & Suriano, Francesco. 2019. KIParla corpus: a new resource for spoken Italian. In Bernardi, Raffaella & Navigli, Roberto & Semeraro, Giovanni (eds.), *Proceedings of the Sixth Italian Conference on Computational Linguistics (CLIC-it 2019)*. CEUR-WS, vol. 2481 (<http://ceur-ws.org/Vol2481/paper45.pdf>) (Accessed on 10.01.2023).
- Meyer, David & Zeileis, Achim & Hornik, Kurt. 2006. The Strucplot Framework: Visualizing Multi-Way Contingency Tables With Vcd. *Journal of Statistical Software* 17(3). 1–48.
- Musi, Elena. 2015. *Dalle apparenze alle inferenze: i predicati sembrare e apparire come indicatori argomentativi*. Lugano: Università della Svizzera Italiana. (PhD dissertation).
- Musi, Elena & Rocci, Andrea. 2017. Evidently epistential adverbs are argumentative indicators: A corpus-based study. *Argument and Computation* 8(2). 175–192.
- Newman, John. 1981. *The semantics of raising constructions*. San Diego: University of California. (PhD dissertation).
- Nissim, Malvina. 2003. Annotation Scheme for Information Status in Dialogue. (Manuscript, University of Edinburgh.)
- Nissim, Malvina & Dingare, Shipra & Carletta, Jean & Steedman, Mark. 2004. An Annotation Scheme for Information Status in Dialogue. In Lino, Maria Teresa & Xavier, Maria Francisca & Ferreira, Fátima & Costa, Rute & Silva, Raquel (eds.) *Proceedings of the 4th Conference on Language Resources and Evaluation (LREC'04)*. (<https://aclanthology.org/L04-1402/>) (Accessed on 10.01.2023)
- Nuyts, Jan. 2001. *Epistemic Modality, Language, and Conceptualization*. Amsterdam-Philadelphia: John Benjamins.
- Partee, Barbara H. 1971. On the requirement that transformations preserve meaning, in: eds. Fillmore, Charles J. & Lagendoen, D. Terence (eds.), *Studies in Linguistic Semantics*. 1–21. New York: Holt, Rinehart and Winston.
- Perek, Florent. 2015. *Argument Structure in usage-based Construction Grammar*. Amsterdam-Philadelphia: John Benjamins.

- Pietrandrea, Paola. 2005. *Epistemic Modality. Functional properties and the Italian system*. Amsterdam-Philadelphia: John Benjamins.
- Polinsky, Maria. 2013. Raising and control. In den Dikken, Marcel (ed.), *The Cambridge Handbook of Generative Syntax*, 577–606. Cambridge: Cambridge University Press.
- Postal, Paul M. 1974. *On Raising: One rule of English grammar and its theoretical implications*. Cambridge: MIT Press.
- Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Cole, Peter (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.
- Prince, Ellen F. 1992. The ZPG Letter. Subjects, Definiteness, and Information-status. In Mann, William C. & Thompson, Sandra A. (eds.), *Discourse Description: Diverse linguistic analyses of a fund-raising text*, 295–326. Amsterdam-Philadelphia: John Benjamins.
- Ransom, Evelyn N. 1977. Definiteness, Animacy, and NP Ordering. In Whistler, Kenneth & Van Valin, Robert D. & Chiarello, Chris & Jaeger, Jeri J. & Petruck, Miriam & Thompson, Henry & Javkin, Ronya & Woodbury, Anthony (eds.), *Proceedings of the 3rd Annual Meeting of the Berkeley Linguistics Society, Berkeley, 19-21 February 1977*, 418–429. Berkeley: Berkeley Linguistic Society.
- Rosenbaum, Peter S. 1967. *The grammar of English predicate complement constructions*. Cambridge: MIT Press.
- Ross, John R. 1967. *Constraints on variables in syntax*. Cambridge: Massachusetts Institute of Technology. (PhD dissertation.)
- Rossini Favretti, Rema & Tamburini, Fabio & e De Santis, Cristiana. 2002. CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In Wilson, Andrew & Rayson, Paul & McEnery, Tony (eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World*, 27–38. München: Lincom-Europa.
- Rubba, Johanna. 1996. Alternate Grounds in the Interpretation of Deictic Expressions. In Fauconnier, Gilles & Sweetser, Eve (eds.), *Spaces, Worlds and Grammar*, 228–261. Chicago-London: University of Chicago Press.
- Ruppenhofer, Josef K. 2004. *The interaction of valence and information structure*. Berkeley: University of California. (PhD dissertation.)
- Salvi, Giampaolo. 1991. I complementi predicativi. In Renzi, Lorenzo & Salvi, Giampaolo & Cardinaletti, Anna (eds.), *Grande Grammatica italiana di consultazione, vol. 2: I sintagmi verbale, aggettivale, avverbiale. La subordinazione*, 191–226. Bologna: Il Mulino.
- Sansò, Andrea. 2006. ‘Agent defocusing’ revisited. Passive and impersonal constructions in some European languages. In Abraham, Werner & Leisiö, Larisa (eds.), *Passivization and Typology: Form and Function*, 232–273. Amsterdam-Philadelphia: John Benjamins.
- Skytte, Gunver & Salvi, Giampaolo. 1991. Frasi subordinate all’infinito. In Renzi, Lorenzo & Salvi, Giampaolo & Cardinaletti, Anna (eds.), *Grande Grammatica italiana di consultazione, vol. 2: I sintagmi verbale, aggettivale, avverbiale. La subordinazione*, 483–570. Bologna: Il Mulino.
- Snider, Neal & Zaenen, Annie. 2006. Animacy and Syntactic Structure: Fronted NPs in English. In Butt, Miriam & Dalrymple, Mary & King, Tracy Holloway (eds.), *Intelligent Linguistic Architectures: Variations on themes by Ronald M. Kaplan*, 323–338. Stanford: CSLI Publications.

- Stefanowitsch, Anatol. 2003. Constructional semantics as a limit to grammatical alternation: The two genitives of English. In Rohdenburg, Günter & Mondorf, Britta (eds.), *Determinants of Grammatical variation in English*, 413–443. Berlin-Boston: De Gruyter Mouton.
- Stefanowitsch, Anatol. 2020. *Corpus linguistics: A guide to the methodology*. Berlin: Language Science Press.
- Squartini, Mario. 2009. Evidentiality, epistemicity, and their diachronic connections to non-factuality. In Hansen, Maj-Britt M. & Visconti, Jacqueline (eds.), *Current Trends in Diachronic Semantics and Pragmatics*, 211–226. Bingley: Emerald Group Publishing.
- Squartini, Mario. 2018. Extragrammatical Expression of Information Source. In Alexandra Y. Aikhenvald (ed.), *The Oxford Handbook of Evidentiality*, 273–286. Oxford: Oxford University Press.
- Szmrecsányi, Benedikt M. 2004. On operationalizing syntactic complexity. In Purnelle, Gérald & Fairon, Cédric & Dister, Anne (eds.), *Proceedings of the 7th International Conference on Textual Data Statistical Analysis, Louvain-la-Neuve, 10-12 mars 2004*, vol. 2, 1032–1039. Louvain: Presses Universitaires de Louvain.
- Thompson, Sandra A. 1990. Information flow and “Dative Shift” in English. In Edmondson, Jerrold & Feagin, Katherine & Mühlhäusler, Peter (eds.), *Development and Diversity: Linguistic Variation across Time and Space (A Festschrift for Charles-James N. Bailey)*, 239–253. Dallas: Summer Institute of Linguistics.
- Uhrig, Peter. 2015. Why the Principle of No Synonymy is Overrated. *Zeitschrift für Anglistik und Amerikanistik* 63(3). 323–337.
- Ungerer, Tobias. forthcoming. Vertical and horizontal links in constructional networks: Two sides of the same coin? *Constructions and Frames*. (<https://tungerer.github.io/files/Ungerer-forthc-Vertical-and-horizontal-links.pdf>) (Accessed on: 13.11.2023)
- Van Valin, Robert D. 1995. Toward a functionalist account of so-called ‘extraction constraints’. In Devriendt, Betty & Goossens, Louis & van der Auwera, Johan (eds.), *Complex Structures: A Functionalist Perspective*, 29–60. Berlin: Mouton de Gruyter.
- von Heusinger, Klaus & Schumacher, Petra B. 2019. Discourse prominence: Definition and application. *Journal of Pragmatics* 154. 117–127.
- Wasow, Thomas. 1997. Remarks on grammatical weight. *Language variation and change* 9(1). 81–105.
- Yamamoto, Mutsumi. 1999. *Animacy and Reference. A cognitive approach to corpus linguistics*. Amsterdam-Philadelphia: John Benjamins.