Silvia Ballarè and Massimo Cerruti

Sociolinguistic variation, or lack thereof, in the use of the Italian subjunctive: mood selection with factive and semi-factive governors

Abstract: This paper deals with various aspects of the use of the subjunctive in spoken Italian. Firstly, a brief overview is provided of the interplay of linguistic and extra-linguistic factors underlying the selection of the subjunctive in various syntactic structures. Special attention is paid to the differences in subjunctive use in factual and non-factual contexts, pointing out that the former are less studied and less affected by normative injunctions than the latter. The focus then falls on factual contexts, in particular subordinate clauses which depend on factive and semi-factive governors. The investigation is based on KIParla, a corpus of spoken Italian consisting of different types of interactions with speakers of various origins and social backgrounds. It is argued that in the factual context at hand the subjunctive fulfills a specific set of functions, mostly related to the speaker's stance towards the content of the subordinate clause. Finally, a statistical analysis is performed using conditional inference trees and random forests in order to uncover the linguistic and extra-linguistic conditioning of subjunctive use. The analysis shows that in the context under observation it is only linguistic factors that significantly condition mood selection; the use of the subjunctive is substantially insensitive to socio-stylistic and geographic variation, and therefore a socially shared pattern of variation has emerged.

Keywords: spoken Italian, sociolinguistic variation, subjunctive, factivity, statistical analysis

1 Introduction

The subjunctive – and its alleged death – often takes center stage in discussions on the "state of health" of standard Italian. This is true not only for linguists but also

Massimo Cerruti, University of Turin, massimosimone.cerruti@unito.it
Silvia Ballarè, University of Bologna, silvia.ballare@unibo.it

for national media, which often devote their attention to it in order to underline the purported corruption of the language. As a matter of fact, the use of the subjunctive is a highly salient feature from a sociolinguistic perspective, as (most) speakers are well aware that in a large number of contexts the subjunctive is to be preferred over the indicative, especially in formal styles. Despite its sociolinguistic saliency, the standard use of the subjunctive is, however, not always straightforward. In fact, not even grammars unambiguously state whether the standard variety requires the use of the subjunctive or of the indicative in specific contexts. In order to achieve a better understanding of subjunctive use, in this paper we analyse the alternation between the subjunctive and the indicative in spoken Italian, with a focus on one type of context which normative grammars usually gloss over, i. e. subordinate clauses depending on factive and semi-factive governors.

The paper opens with an overview of the use of the subjunctive in contemporary Italian, taking into account its semantic properties and its sociolinguistic characterization (Section 2), and it then goes on to present our case study in Section 3. Section 3.1 discusses in some detail the data and methodology on which this paper is based, presenting the main source of data – the KIParla corpus (Mauri et al. 2019) – as well as the criteria behind the selection of our dataset and the linguistic and extralinguistic factors considered in our analysis. After some preliminary observations on the functional properties of the subjunctive (Section 3.2), we present the results of our quantitative analysis on mood selection in the dataset, with a focus on the statistical tests (i. e. inference trees and random forests), which provide information about the relevance of our predictors. As we argue, the results shed light on the oscillation between the two moods: while extra-linguistic factors seem to play almost no role, other linguistic predictors appear to be unexpectedly highly relevant (Section 3.3). In the final paragraph (Section 4), we draw our conclusions[1].

2 Subjunctive use in contemporary Italian

The Italian subjunctive occurs in independent non-declarative clauses and subordinate clauses. In most of these contexts, its occurrence is subject to a complex array of semantic, lexical, morphosyntactic, and extra-linguistic factors. Semantic factors

are particularly relevant in non-factual complement clauses, in which the use of the subjunctive is triggered by matrix predicates "expressing volition and deontic necessity, modulating the epistemic commitment, or evaluating the propositional content" (Squartini 2010: 245); one such case is the use of the subjunctive in clauses depending on opinion verbs, as in (1). However, the subjunctive is found to occur even in complement clauses governed by assertive verbs, as in (2); an evidential interpretation has been proposed for such cases (Renzi 2019: 23–25), as speakers would explicitly distance themselves from the content of the clause by stressing their role as the non-source (cf. Squartini 2008: 939).

(1)   *Credo che adesso si **apra** una fase diversa, una fase nuova*
      'I think that now a different phase, a new phase, is opening'
      (LIP Corpus; cf. Schneider 1999: 87)
(2)   *la Valletta sostiene che gli interventi di salvataggio in mare **siano** avvenuti nella zona SAR (Ricerca e Soccorso) libica*
      'Valletta claims that sea rescue operations have been carried out in the Libyan SAR (Search And Rescue) area'
      (*La Stampa*, 13.05.2018; Renzi 2019: 21)

In most syntactic structures, the subjunctive is found to alternate with the indicative. In some cases, the alternation between the subjunctive and the indicative is determined by the semantic features of one or more components of the utterance; for instance, the referring head of a factual relative clause triggers the indicative (e. g. *cerco il ragazzo che voleva cantare*, 'I'm looking for the boy who wanted to sing'), while the non-referring head of a non-factual relative clause triggers the subjunctive (e. g. *cerco un ragazzo che voglia cantare*, 'I'm looking for a boy who wants to sing'). In other cases, the alternation between the subjunctive and the indicative determines a different shade in meaning (cf. Mauri and Sansò 2016: 172); for example, the use of the indicative in lieu of the subjunctive would indicate a higher degree of epistemic commitment in (1) and would not overmark any evidential distance in (2) (see also Schneider 1999: 136–138; Prandi 2012: 113).

Nevertheless, the alternation between the subjunctive and the indicative is not always semantically motivated. In fact, there is evidence to suggest that the subjunctive is undergoing a process of grammaticalization (cf. Bybee et al. 1994: 214), which is currently reflected in "lexical routinization […] and structural conventionalization" (Poplack et al. 2018: 217). A recent variationist study on complement clauses based on data from LIP and C-ORAL-ROM corpora (Digesto 2019, 2021) indeed shows that the selection of the subjunctive is largely determined by lexical and morphosyntactic factors. In particular, distributional and statistical evidence reveals that subjunctive selection is mainly promoted by a set of lexical governors

distributed across various semantic classes, as well as by the embedded suppletive forms of *essere*; a significant effect on the use of the subjunctive is also contributed by complementizer deletion and non-affirmative sentences (cf. Digesto 2019: 103–148). Studies adopting other approaches and/or relying on other data sets show that subjunctive selection may also be favored by further syntactic factors; one of the most notable is the thematic function, and in particular the left-dislocated position, of the complement clause (cf. Wandruszka 1991: 418–420, 454; Schneider 1999: 20, 58–59, 152, 195–196; Blücher 2003: 172–173; Prandi 2012: 102; Squartini 2010: 245).

However, the alternation between the subjunctive and the indicative is promoted by both internal and extra-linguistic factors, the latter being responsible for the identification of language varieties. In a large number of cases, in fact, the selection of either the subjunctive or the indicative is associated with the use of different varieties of Italian. In particular, it often happens that the selection of the subjunctive in a certain syntactic structure is proper to standard Italian, and, therefore, imbued with linguistic prestige, while the selection of the indicative in the same structure is typical of sub-standard varieties, such as informal styles and uneducated speech (cf., *inter alia*, Voghera 1993; Schneider 1999; Lombardi Vallauri 2003; Prandi 2012; Poplack et al. 2018)[2]. Below are some examples of the use of the indicative, as opposed to the subjunctive, in non-factual relative clauses (3), in relative clauses with a superlative antecedent (4), and in the protasis of non-factual conditional clauses (5). Yet, in these and other cases as well, the selection of the indicative is also accepted in neo-standard Italian (cf. Berruto 2012: 80–81; Ballarè 2020: 478–482).

(3)   *tutti e due li puoi fare ma* [...] *deve essere un lavoro che ti **impegna** quelle due tre ore al giorno*
       'you can do both but [...] it has to be a job that takes you two or three hours a day'
       (KIParla Corpus, PTD006)

(4)   *il viaggio più bello che **ho fatto** [...] è quello che ho fatto nel 2017 in Sardegna*
       'the most beautiful trip I've had [...] is the one I had in 2017 in Sardinia'
       (KIParla Corpus, PTD006)

(5)   *se lo **beccavano**, ammazzavano anche gli altri*
       'if they had found him, they would have killed the others, too'
       (KIParla Corpus, PTB014)

---

2 It is no coincidence that indicatives "are [...] (i) structurally less complex, (ii) show fewer distributional constraints, and (iii) show a higher number of inflectional possibilities than subjunctives" (Mauri and Sansò 2016: 176).

At the same time, it can happen that the subjunctive is used in syntactic structures in which standard Italian requires the indicative. In such cases, the use of the subjunctive is generally characteristic of 'supra-standard' varieties (in the sense of Cerruti 2017), i. e. language varieties higher than standard in the *Architektur der Sprache*, such as refined formal and educated varieties[3]. Examples include the selection of the subjunctive in concessive clauses introduced by *anche se*, as in (6) (cf. Mazzoleni 1991: 794–795; Renzi 2019: 27), and in factual relative clauses, as in (7) (cf. Santulli 2009: 174; Gualdo 2014: 235–236). Supra-standard features generally enjoy high prestige, while sub-standard features are often affected by negative attitudes and stigmatization. Utterances (3) and (7) also show that sub-standard varieties can opt for the indicative and supra-standard varieties for the subjunctive even in syntactic structures in which the selection of either the indicative or the subjunctive is meaningful in standard Italian (such as in factual and non-factual relative clauses, as noted above).

(6)   *anche se l'accordo* **abbia** *attualmente pochissimo valore*
      'even if the agreement is currently of very little value'
      (RAI Radio Tre, 25.09.2018; Renzi 2019: 27)
(7)   *non bisogna cullarsi nell'illusione che la lentezza con la quale si* **stia** *procedendo possa essere indicativa di un fallimento futuro*
      'one must not be lulled into the illusion that the slow progress being made is indicative of future failure'
      (NUNC Corpus; cf. Santulli 2009: 174)

Moreover, the use of the indicative in lieu of the subjunctive is traditionally seen as more frequent in Central and Southern varieties of Italian (cf. Rohlfs 1969: 59; Maiden 1995: 258; Schneider 1999: 188; Berruto 2012: 80–81), also because of the influence of their corresponding Italo-Romance dialects. However, the results of the latest available study on subjunctive selection (Digesto 2019) are actually at odds with such supposed differences between Central/Southern and Northern varieties. In fact, the comparison of parallel multivariate analyses across the four geographically-distinct datasets of LIP and C-ORAL-ROM corpora – Milan, Florence, Rome, and Naples – shows that the use of the subjunctive has the same pattern of variability across the four cities (Digesto 2019: 147–154; similar results are found in Lombardi Vallauri 2003)[4].

---

3 In particular, see Berruto's (2012: 24) multidimensional model of the 'architecture' of Italian.
4 As for substratum influence, Benincà (1993: 262) points out that the same speakers who regularly use the subjunctive in Northern Italo-Romance dialects tend to use the indicative in Northern varieties of Italian, especially in colloquial speech.

Finally, the bulk of studies have focused on the selection of the indicative in lieu of the subjunctive in contexts encoding non-factual events. In most of these contexts, the use of the indicative is often proscribed by normative grammars but is currently extending its reach from sub-standard varieties, such as colloquial speech and informal styles, to neo-standard Italian (see above), thus resembling the diffusion of a 'change from below' (cf. Labov 1994: 78). This is, however, the development of a pattern of variability that was already operative in the past, as the alternation between the subjunctive and the indicative in non-factual contexts can consistently be found from Old Italian onward (Berruto 2012: 80). Less attention has been paid to mood selection – and in particular to the use of the subjunctive in lieu of the indicative – in contexts encoding factual events. Normative injunctions are not always in agreement regarding mood selection in such contexts, and several grammars do not even address this issue (see, e. g., Stewart 2002). Scholars themselves have different opinions on which behavior is to be deemed standard for factual states of affairs (cf. Prandi 2012: 100–103; Renzi 2019: 19–28) and/or accept variability in mood selection as the norm (Wandruszka 1991: 472–481). Some note an increasingly generalized use of the subjunctive in factual contexts nowadays and consider it as a 'change from above' (Renzi 2019: 19). The presence of the subjunctive in such contexts is actually found from Old Italian onward, but it has always been very uncommon (and much more frequent in formal written usage) and limited almost exclusively to certain structures (e. g. complement clauses depending on evaluative verbs; cf. Schneider 1999: 24–25).

3 Subjunctive use in factual contexts: a case study

In what follows, the focus will fall on the use of the subjunctive in contexts encoding factual events. Special attention will be paid to the selection of the subjunctive in a specific factual context, i. e. that of subordinate clauses depending on factive and 'semi-factive'[5] governors (be they verbs, nouns, adjectives, or adverbs; see, *inter alia*, Kiparsky and Kiparsky 1971; Hooper and Thompson 1973; Wandruszka 1991: 472–481, 648; Schulz 2003: 6–112).

---

5 Here we follow Hooper and Thompson (1973) in considering semi-factives the governors that presuppose the truth of their complements only in some environments.

3.1 Data and methods

Our analysis is based on the two modules that currently constitute the KIParla corpus (cf. Mauri et al. 2019, www.kiparla.it), i. e. KIP and ParlaTO. The KIP module consists of approximately 70 hours (661.175 tokens) of speech data collected in Turin and Bologna, and it includes various types of interactions recorded in the academic setting, both between students, i. e. semi-structured interviews and spontaneous conversations, and between students and professors, i. e. office hours, oral examinations, and lectures. The ParlaTO module consists of approximately 50 hours (552.461 tokens) of speech data collected in Turin, and it includes semi-structured interviews with speakers from different social backgrounds, e. g. people of various ages (ranging from 18 to over 85) with different levels of education (i. e. elementary/middle school, technical/vocational school, high school, BA/MA, PhD; cf. Cerruti and Ballarè 2020). Moreover, both KIP and ParlaTO embrace a great variety of speaker provenance; overall, the speakers involved in the recordings come from all twenty regions of Italy, as well as from foreign countries (cf. Mauri et al. 2022).

We have structured the dataset as follows. Firstly, drawing on both previous literature (Schmitt Jensen 1970; Kiparsky and Kiparsky 1971; Wandruska 1991: 472–481; Schneider 1999; Schulz 2003) and reference grammars (Wandruska 1991: 472–481, Serianni 2010 [1989], ch. XIV, §§ 29 et seq.), we singled out more than 300 factive and semi-factive governors. Next, we performed a systematic search of the KIParla corpus to identify the set of governors followed by the subordination marker *che*, and only 64 factive and semi-factive governors (out of 300) were found to occur. Finally, we manually selected the cases in which these 64 governors were actually used with a factive meaning, i. e. presupposing the truth of their complements. The definitive dataset includes 711 occurrences, produced by 176 speakers.

Afterwards, we annotated each occurrence according to the mood of the verb (*subjunctive* or *indicative*) in the subordinate clause and to 11 internal and external factors that might be relevant for explaining the mood selection.

Internal factors:
1) PoS of the governor
   3 levels: *adjective* or *adverb*[6] (e. g. *sono contenta che* …, "I am happy that …", *è bene che* …, "it is good that …"), *noun* (e. g. *il fatto che* …, "the fact that …"), *verb* (e. g. *so che* …, "I know that …");

---

6 Adjectives and adverbs have been categorized together because they share syntactic properties in terms of distribution and modification (especially in the contexts under scrutiny).

2) Topic (thematic function of the clause)
   2 levels: *yes* (thematic clause), *no* (the subordinate clause is included in the asserted part of the utterance);
3) Distance between *che* ("that") and V2 (i. e. the verb of the subordinate clause)
   2 levels: *adjacent* (*che* and the verb are adjacent or separated by a negative morpheme or a clitic pronoun), *non-adjacent* (all other cases);
4) Negation
   2 levels: *yes* (presence of a negative marker in the main clause), *no* (all other cases);
5) V2 *essere*
   2 levels: *essere* (all cases in which the verb "to be" is inflected), *other* (all other cases);
6) Tense (of V2)
   2 levels: *present*, *past*;

External factors:
7) Formality
   2 levels: *formal* (interactions in which the participants have a highly asymmetrical relationship, i. e. lessons, oral exams, and office hours), *informal* (contexts in which the speakers have a more symmetrical relationship, i. e. semi-structured interviews and spontaneous conversations);
8) Speaker's gender
   2 levels: *female*, *male*;
9) Speaker's age range
   3 levels: *youth* (30 and under), *adults* (between 31 and 60), *seniors* (61 and over);
10) Speaker's geographic origin
    4 levels: *north*, *center*, *south and islands, abroad*;
11) Speaker's educational level
    2 levels: *lower* (elementary school diploma, middle school diploma, high school diploma *and* currently working and not studying), *higher* (university students, BA or MA degree, PhD).

After the annotation process, we decided to run two statistical tests in order to evaluate the importance of the factors being considered for mood selection. Given the structure of our dataset (i. e. 711 occurrences, 1 independent variable, and 11 predictors), we opted for conditional inference trees and random forests; in fact, they are highly recommended in cases in which there are few linguistic occurrences and many predictors and, above all, they can work with unbalanced data (for a thorough discussion, see Tagliamonte and Baayen 2012 and Levshina 2015: 291–300).

In order to construct a conditional inference tree, the algorithm considers all the predictors and "splits" the data into subsets that are increasingly homogeneous. The result provides a visualization of the data distribution according to the relevance of predictors that have proved to be useful in creating homogeneous data subsets.

Random forests consist of a number of conditional inference trees, and thus they can be considered to be more sophisticated and reliable. The output of the analysis is a ranking of the importance of each predictor. However, it does not provide the "direction" in which the predictor plays a role; that is to say, it can highlight, for example, the relevance of the distance between *che* and V2, but it does not indicate whether the adjacency triggers the selection of the subjunctive or the indicative. To understand this aspect, it is crucial to consider the distribution of the scrutinized feature in the given context.

Some preliminary remarks will be made in the next section. The focus will then turn to the results of the statistical tests.

3.2 Preliminary remarks

In our dataset, the subjunctive is used in 16.88 % of cases (N=120/711) and by 39.20 % of speakers (N=69/176).

First of all, some preliminary observations can be made regarding the functional properties of the subjunctive. Our data suggest that in most subordinate clauses depending on factive and semi-factive governors, the subjunctive fulfills a set of functions that can be subsumed under a broader category. In the context under scrutiny, in fact, the subjunctive appears to serve functions which are mostly "connected with the speaker's 'stance' or attitude towards the propositional content" (Mauri and Sansò 2016: 174). In particular, in the vast majority of cases, the subjunctive encodes a factual state of affairs that is subject to the speaker's evaluation and/or is at odds with the speaker's expectations (see the notion of 'mirativity', DeLancey 1997), as demonstrated, for instance, in (8) and (9) below. In the former, the subjunctive is used in a complement clause governed by an evaluative predicate; in the latter, the subjunctive encodes information which is (presented as) unexpected to the 'speaker' (i. e. the subject of the matrix clause). Moreover, in both (8) and (9), the content of the complement clause is thematic and given. In such cases, the content of the complement clause is put in the background, while the content of the matrix clause – i. e. the speaker's stance – is foregrounded (cf. Wandruszka 1991: 419–420; Schneider 1999: 156–158; Squartini 2010: 245).

(8) *non dovrebbe essere Presidente del Consiglio, però è bello che lo **sia***
'he should not be Prime Minister, but it's nice that he is'
(KIParla Corpus, BOA3021)

(9) *si guarda allo specchio, poi dice ti guardi il naso che ti pende a destra, lui non si era*
*mai accorto che il suo naso **pendesse** a destra, e comincia a pensare al suo naso*
che pende a destra
'he looks at himself in the mirror, then tells himself you're staring at your nose
which hangs to the right, he never noticed that his nose hangs to the right, and
begins to think about his nose hanging to the right'
(KIParla Corpus, BOD1006)

Furthermore, in subordinate clauses depending on factive and semi-factive governors, the use of the subjunctive can sometimes be seen as reinforced by evidentiality (cf. Section 2). In (10), for example, the subjunctive form *sappiano* occurs in a
thematic clause whose informational content is subject to the speaker's evaluation
and is based on an external source of evidence.

(10) TO073: *mi hanno detto* [...] *di parlare direttamente col professore* [...] *per capire*
se fosse possibile partire [...]
'(the administrative offices) told me [...] to speak directly with the professor [...] to see if it were possible to leave [...]'
 TO074: *ma non dipende da me, dipende dagli uffici* [...] *è strano che non glielo*
***sappiano** dire*
'but it's not up to me, it's up to the administrative offices [...] it's strange
that they can't tell you (if it's possible to leave)'
(KIParla Corpus, TOA1004)

In a few cases, however, the subjunctive is drawn on in syntactic structures which
can hardly, if at all, be associated with the speaker's stance. This is often the case
with subordinate clauses governed by *il fatto che* ('the fact that'), the meaning of
the latter being unrelated to the adoption of a stance, as in utterance (11). Here,
the subjunctive encodes a factual state of affairs that is not subject to the speaker's
evaluation, nor is it presented as something unexpected to the speaker.

(11) *il fatto che adesso, in qualunque museo, voi **andiate** col cellulare sta cambiando*
l'esperienza museale
'the fact that now in any museum you can bring your mobile phone is changing
the museum experience'
(KIParla Corpus, TOD1002)

3.3 Trees and forests

In this section we present the results of a statistical analysis carried out using conditional inference trees and random forests. Figure 1 illustrates the output of the conditional inference tree. The C-index of the model is 0.9, thus indicating "outstanding discrimination" (Hosmer and Lemeshow 2000: 162).
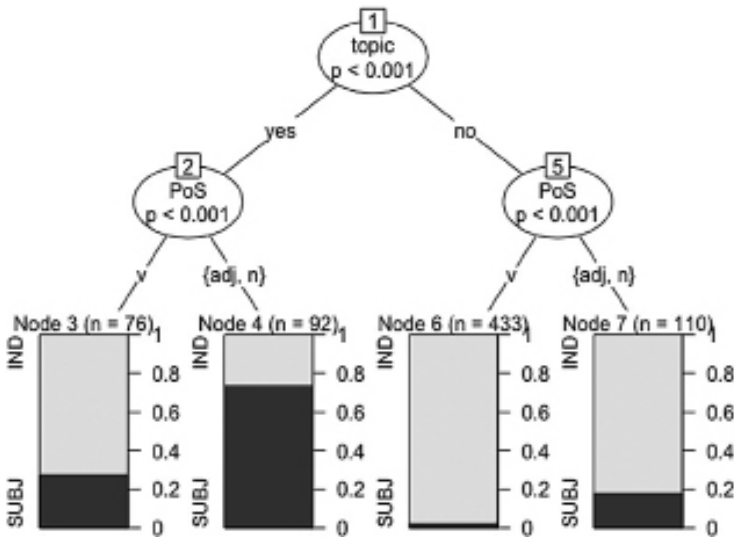

Figure 1: Inference tree

Figure 1 shows that the algorithm selected two predictors to be relevant in creating a homogeneous data subset, i. e. Topic and PoS of the governor (cf. Section 3.1). The first point to note is that if the subordinate has thematic content, then V2 will be more frequently inflected in the subjunctive. For the sake of clarity, we report the mood distribution according to this predictor in Table 1.

Table 1: Mood distribution

Topic	subj.	ind.	Tot.
Yes	89 (53 %)	79 (47 %)	168 (100 %)
No	31 (5.7 %)	512 (94.3 %)	543 (100 %)
Tot	120	591	711
			Chi square test – P value < 0.00001

The first line of Table 1 indicates that the subjunctive is selected in more than half of the occurrences (53 %), as exemplified in (12). The use of the subjunctive in this factual context is attested in the literature (see Squartini 2010: 145 et seq., Wandruszka 1991: 418–420 *inter al.*), and it can be explained by the fact that the thematic content is not asserted – and thus does not necessarily require the indicative – but rather is pragmatically presupposed (or, more precisely, previously activated in discourse).

(12) BO145: *e la critica, cioè, c'è sempre stata e questa non la possiamo togliere*
'and critics, I mean, they have always been there and we cannot dismiss them'
BO139: *certo*
'sure'
BOI145: *ed è bello che ci **sia** la critica, perché è costruttiva*
'and it is good that critics exist, because they are constructive'
(KIParla Corpus, BOA3018)

The second line of Table 1 indicates that the subjunctive is attested in very few cases (5.7 %); in fact, it occurs only rarely in a rhematic clause, in which the propositional content is neither stated in previous discourse nor (presented as) known to the hearer, as in (13).

(13) TOR001: *vedi altri cambiamenti oltre a questo?*
'do you see any other changes besides this one?'
TOI014: *mh, mi rendo conto che l'università a Torino **sia** decisamente una scelta che molte persone optano*
'um, I notice that going to university in Turin is definitely a choice that many people opt for'
(KIParla Corpus, PTA012)

The second most important predictor is the PoS of the governor. As shown in Table 2, the results are more heterogeneous when considering adjectives/adverbs and nouns. If the governor is an adjective/adverb, in more than 60 % of the cases the subjunctive is used; conversely, if the governor is a noun[7], then V2 is inflected in the indicative mood in more than 60 % of the cases. The results are very different when a verb is involved; in fact, if the governor is a verb, then in the vast majority of cases (93.7 %), V2 is inflected in the indicative mood.

---

7 Note that in 162 cases out of 164, the noun is *fatto*, "fact".

Table 2: PoS distribution

PoS	subj.	ind.	Tot.
Adjective/Adverb	31 (62 %)	19 (38 %)	50 (100 %)
Noun	57 (37.5 %)	95 (62.5 %)	152 (100 %)
Verb	32 (6.3 %)	477 (93.7 %)	509 (100 %)
Tot.	89	114	711

The results of the test indicate that there is a macro-distinction in the mood selection of the V2 that depends on the governor. As mentioned before, however, in the literature (see Poplack et al. 2018: 217, Digesto 2021) it has been noted that mood selection can be lexically motivated, in that the use of the subjunctive is routinized with some governors. In order to discuss the possible relevance of lexical factors in our dataset, in Table 3 we provide the mood distribution for the 10 most frequent lexical items per PoS.

Table 3: Lexicon and mood selection

Adjectives/Adverbs			Nouns			Verbs		
lexeme	mood subj.	ind.	lexeme	mood subj.	ind.	lexeme	mood subj.	ind.
Giusto ('right')	8	0	Fatto ('fact')	57	93	Sapere ('know')	7	149
Bene ('good')	2	4	Fastidio ('bother')	0	1	Ricordarsi ('remember')	0	82
Contento ('happy')	2	4	Fortuna ('luck')	0	1	Capire ('understand')	4	52
Normale ('normal')	3	2				Rendersi conto ('realize')	2	29
Strano ('strange')	4	1				Scoprire ('discover')	0	26
Bello ('beautiful')	4	0				Accorgersi ('notice')	1	23
Inutile ('useless')	1	3				Dovere ('must')	0	21
Meglio ('better')	1	3				Notare ('note')	0	14
Importante ('important')	2	1				Sentire ('feel')	1	16
Tanto ('a lot')	2	0				Dimostrare ('prove')	0	9

First of all, the data in Table 3 confirms what we have already observed in the inference tree discussed above. For adjectives/adverbs and nouns, mood distribution is more varied, while for verbs it is more homogeneous. In fact, as far as adjectives/ adverbs and nouns are concerned, in the majority of cases (7 out of 10) variation in mood selection occurs, even when considering one single lexical element (see, for example, *contento* "happy" and *normale* "normal"). With 2 governors (*giusto* "right" and *tanto* "a lot") the subjunctive is categorical, and with 4 (*normale* "normal", *strano* "strange", *bello* "beautiful", and *importante* "important") it is more frequent but alternates with the indicative. Finally, with 4 lexical items (*bene* "good", *contento* "happy", *inutile* "useless", and *meglio* "better") the indicative is more frequent.

Instead, when the governor is a verb, the indicative is far more common with (almost) every governor; the subjunctive is present in sporadic cases and is always in competition with the indicative (see, for example, *sapere* "know" and *sentire* "feel"). A few cases, however, which are not reported in the table given their lower absolute frequencies, do not fit into this picture. In fact, three governor verbs are more often followed by the subjunctive than the indicative. Such is the case with *piacere* "like" (4 subj VS 2 ind), *non togliere* "remains" (4 subj VS 2 ind), and *dispiacere* "dislike" (2 subj). Even if the data are limited, these last distributions suggest that lexical factors may actually play a role in the mood selection of the verb, at least in some cases.

Furthermore, if we focus on the semantics of the governors, it emerges that most of the adjectives and adverbs have an evaluative meaning, as exemplified above in (10) and in (14); this could be the reason why they trigger the selection of the subjunctive, since they relate to the speaker's attitude towards the propositional content.

(14) *È giusto che ne **leggiate** almeno alcuni brani*
      'It is right that you read at least some excerpts of it'
      (KIParla Corpus, BOD1007)

The "noun" category is nearly coincident with *fatto*, "fact", since the item is involved in 162 cases out of 164. In these cases, however, the semantics of *fatto* clearly has nothing to do with the speaker's stance, since it is a construction adopted for nominalization and, unlike adjectives/adverbs, does not convey any evaluation. At least here we can explain the use of the subjunctive by considering the informative structure of the utterances (see also Loengarov 2005). In fact, *il fatto che* is often used to introduce the thematic content, as exemplified in (15):

(15) *Sì e poi sicuramente il fatto che Torino non **sia** più una città industriale ha cambiato tutto*
'Yes and then for sure the fact that Turin is no longer an industrial city has changed everything'
(KIParla Corpus, PTD004)

Finally, it is worth noting that verb semantics seems to be linked with evidentiality (see *sapere* "know") and often with mirativity (see *rendersi conto* "realize" and *accorgersi* "notice"). Considering the data, we can say that speakers only rarely employ the subjunctive to encode their expectations, since the indicative is far more frequent with every verb, as in (16). As mentioned above, there are, however, three exceptions: the subjunctive is more frequent than the indicative with 3 verbs, and it must be stressed that 2 of them (i. e. *piacere* "like" and *dispiacere* "dislike", "feel sorry for") clearly have an evaluative meaning, as exemplified in (17).

(16) *Ma io te lo giuro, mi sono resa conto che qua non **andava** bene dal primo giorno di università*
'But I swear, I realized that things were not going well here from the first day of university'
(KIParla Corpus, BOA3003)

(17) *Mi piace appunto che **sia** un quartiere molto vivo*
'I indeed like that it is a very lively neighborhood'
(KIParla Corpus, PTD007)

Summing up, we can say that the data would seem to indicate that the subjunctive more often encodes that the factuality of the clause is subject to the speaker's evaluation, and, very rarely does it encode that the factuality of the clause is at odds with the speaker's expectations.

In relation to the second part of our analysis, Figure 2 illustrates the output of the random forest by mapping the relative importance of the factors under consideration. The C-index for the model is >0.9 and thus demonstrates outstanding discrimination (Hosmer and Lemeshow 2000: 162).

First of all, the random forest confirms the results obtained with the inference tree. In fact, the two predictors that play a role in mood selection are the informative status of the subordinate clause and the PoS of the governor. As shown in Figure 2, the third (i. e. *essere*) parameter is by far less important; its low relevance is also supported by the fact that if we perform the chi-square test on mood distribution according to this parameter, we obtain a high p value ($p = 0.196301$), thereby attesting its insignificance. We can therefore assume that all the other predictors that have a lower ranking do not actually play any role in mood selection. From a sociolinguistic perspective, what is more striking is what is *not* important, also

considering what has been traditionally sustained in the literature[8]. In fact, none of the extralinguistic predictors has proved to be relevant. Surprisingly, at least in our dataset, neither formality nor education, nor even the geographic origin of the speaker has any relevance in the model. In other words, they do not help us in understanding mood distribution in the subordinate clauses analyzed.
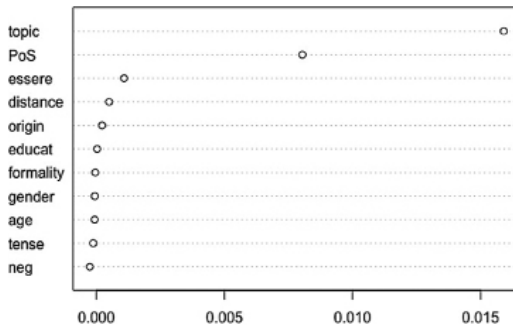

Figure 2: Random forest

4 Conclusions

Previous studies indicate that the use of the Italian subjunctive is subject to a complex array of factors. On the one hand, subjunctive use conveys a wide range of meanings and pragmatic functions. On the other hand, as a result of a process of grammaticalization, the use of the subjunctive is undergoing lexical routinization and structural conventionalization; therefore, it is not always semantically motivated. Moreover, subjunctive selection is associated with the use of different varieties of Italian. In particular, in most syntactic structures the selection of the subjunctive is proper to standard Italian, while the selection of the indicative is characteristic of sub-standard varieties and/or neo-standard Italian. However, this seems to be especially true in non-factual contexts.

As regards factual contexts, our findings suggest that the use of the subjunctive with factive and semi-factive governors has some distinguishing features. In fact, in subordinate clauses dependent on such governors, subjunctive selection appears to fulfill a clearly defined set of functions and seems to be hardly, if at all, conditioned by sociolinguistic variation. On the functional side, the use of the subjunc-

---

8 However, it is important to mention once again that most of the studies have focused their attention on non-factive governors, while factive and semi-factive ones have been rather neglected.

tive signals first and foremost that the content of the clause is thematic (cf. Section 3.3.). This is consistent with the well-known non-etymological (or non-harmonic, cf. Bybee et al. 1994: 218 ff.) function developed by the Romance subjunctive, i. e. that of signaling the theme of the sentence (cf. Lamiroy and De Mulder 2011). Moreover, when the governor conveys the speaker's stance towards a factual state of affairs, this 'thematic' subjunctive helps to put the content of the subordinate clause in the background and, consequently, foregrounds the content of the matrix clause, i. e. the speaker's stance. This is especially the case with factual states of affairs which are subject to the speaker's evaluation (see the case of mood selection with adjectives and adverbs, as well as with verbs having an evaluative meaning; Section 3.3.). These functional properties play such a significant role in subjunctive use that they seem to override lexical routinization and structural conventionalization.

As far as sociolinguistic variation is concerned, none of the extra-linguistic factors taken into account contributes statistically to mood selection. The use of the subjunctive in subordinate clauses depending on factive and semi-factive governors is, therefore, substantially insensitive to socio-stylistic and geographic variation. This result is particularly important if we consider that the use of the subjunctive in the context at hand is scarcely affected by normative pressure, as several grammars fail to address mood selection in factual contexts, and the few that deal with it are often in disagreement (cf. Section 2). A communal norm, i. e. a socially shared pattern of subjunctive selection, has thus developed in the dearth of normative injunctions or despite conflicting injunctions. The use of the subjunctive with factive and semi-factive governors can hence be referred to as "standard by (mere) usage" (Ammon 2003: 2).

Finally, the emergence of such a communal norm can be seen as an outcome of the increasing use of the subjunctive in factual contexts, which some scholars consider to be a 'change from above' (cf. Section 2). Conceiving the spread of the subjunctive in factual contexts as a conscious change (i. e. a change coming from above the level of the speaker's awareness) is actually problematic. In fact, given that a (consistent) benchmark for mood selection in factual contexts is not established, speakers are not necessarily aware that the use of the subjunctive can introduce a change. However, the traditionally high status of the subjunctive corroborates the idea that subjunctive use in factual contexts has originated in language varieties enjoying high prestige, such as refined formal and educated varieties. Moreover, speakers are more likely aware that in a large number of contexts the use of the subjunctive is imbued with linguistic prestige, and this very awareness could contribute, or could have contributed, to increasing the use of the subjunctive. In the end, it can be argued that subjunctive selection in factual contexts has actually extended its reach beyond refined formal and educated varieties and, at least in subordinate clauses which depend on factive and semi-factive governors, has come to be subject to a socially shared pattern of variation.

References

Ammon, Ulrich. 2003. On the social forces that determine what is standard in a language and on conditions of successful implementation. *Sociolinguistica* 17. 1–10.

Ballarè, Silvia. 2020, L'italiano neo-standard oggi: stato dell'arte. *Italiano LinguaDue* 12(2). 469–492.

Benincà, Paola. 1993. Sintassi. In Alberto A. Sobrero (ed.), *L'italiano: le strutture, la variazione*, I, 247–290. Bari: Laterza.

Berruto, Gaetano. 2012. *Sociolinguistica dell'italiano contemporaneo*. Roma: Carocci.

Blücher, Kolbjörn. 2003. Modalità, modo, 'concordanza modale'. Una prospettiva teorica. In Mathée Giacomo-Marcellesi and Alvaro Rocchetti (eds.), *Il verbo italiano. Studi diacronici, sincronici, contrastivi, didattici. Atti del XXXV Congresso Internazionale della Società di linguistica italiana (Parigi, 20–22 settembre 2001)*, 169–177. Roma: Bulzoni.

Bybee, Joan, Revere Perkins, and William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.

Cerruti, Massimo. 2017. Changes from below, changes from above: relative constructions in contemporary Italian. In Massimo Massimo, Claudia Crocco and Stefania Marzo (eds.), *Towards a New Standard*, 62–88. Berlin/New York: Mouton De Gruyter.

Cerruti, Massimo and Silvia Ballarè. 2020. ParlaTO: corpus del parlato di Torino. *Bollettino dell'Atlante Linguistico Italiano (BALI)* 44. 171–196.

DeLancey, Scott. 1997. Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1(1). 33–52.

Digesto, Salvio. 2019. *A variationist analysis of subjunctive variability across space and time: from contemporary Italian back to Latin.* Ottawa: University of Ottawa PhD dissertation.

Digesto, Salvio. 2021. Lexicalization and social meaning of the Italian subjunctive. *Cadernos del Linguistíca* 2(3). https://cadernos.abralin.org/index.php/cadernos/article/view/609

Gualdo, Riccardo. 2014. Il "parlar pensato" e la grammatica dei nuovi italiani. *Studi di grammatica italiana* XXXIII. 223–254.

Hooper, Joan B. and Sandra. A. Thompson. 1973. On the applicability of root transformations. *Linguistic Inquiry* 4. 465–498.

Hosmer, David W. and Stanley Lemeshow. 2000. *Applied logistic regression*. New York: Wiley.

Kiparsky, Paul and Carol Kiparsky. 1971. Fact. In Manfred Bierwisch and Karl Erich Heidolph (eds.), *Progress in Linguistics*, 143–173. Berlin: Mouton de Gruyter.

Labov, William. 1994. The study of change in progress: Observations in real time. In William Labov, *Principles of linguistic change: Internal factors*, 73–112. Blackwell: Oxford.

Lamiroy, Béatrice and Walter De Mulder. 2011. Degrees of grammaticalization across languages. In Bernd Heine and Heiko Narrog (eds.), *The Oxford Handbook of Grammaticalization*, 302–318. Oxford: Oxford University Press.

Levshina, Natalia. 2015. *How to do Linguistics with R*. Amsterdam: John Benjamins.

Loengarov, Alexander. 2005. Le fait que? et la question du subjonctif: la directionnalité de la grammaticalisation. *Cahier Chronos* 12. 67–82.

Lombardi Vallauri, Edoardo. 2003. Vitalità del congiuntivo nell'italiano parlato. In Teresa Poggi Salani and Nicoletta Maraschio (eds.), *Italia linguistica anno Mille, Italia linguistica anno Duemila. Atti del XXXIV Congresso Internazionale della Società di linguistica italiana (Firenze 19–21 ottobre 2000)*, 609–635. Roma: Bulzoni.

Maiden, Martin. 1995. *A Linguistic History of Italian*. London/New York: Longman.

Mauri, Caterina and Andrea Sansò. 2016. The linguistic marking of (ir)realis and subjunctive. In Jan
      Nuyts and Joan van der Auwera J. (eds.), *Handbook of modality and mood*, 166–195. Oxford: Oxford
      University Press.
Mauri, Caterina, Silvia Ballarè, Eugenio Goria, Massimo Cerruti and Francesco Suriano. 2019. KIParla
      corpus: a new resource for spoken Italian. In Raffaella Bernardi, Roberto Navigli and Giovanni
      Semeraro (eds.), *Proceedings of the 6th Italian conference on Computational Linguistics CLiC-it*,
      Torino, Accademia University Press. http://ceur-ws.org/Vol-2481/paper45.pdf
Mauri, Caterina, Silvia Ballarè, Eugenio Goria and Massimo Cerruti. 2022. Il corpus KIParla. In *Atti del
      LIV Congresso Internazionale della SLI*, 109–118. Milano: Officinaventuno.
Mazzoleni, Marco. 1991. Le frasi ipotetiche. In Lorenzo Renzi and Giampaolo Salvi (eds.), *Grande
      grammatica italiana di consultazione II*, 751–817. Bologna: Il Mulino.
Poplack, Shana, Rena Torres Cacoullos, Nathalie Dion, Rosane de Andrade Berlinck, Salvio Digesto,
      Dora Lacasse, and Jonathan Steuck. 2018. Variation and Grammaticalization in Romance: A
      Cross-Linguistic Study of the Subjunctive. In Wendy Ayres-Bennett and Janice Carruthers (eds.),
      Manuals in Linguistics: Romance Sociolinguistics, 217–252. Berlin/New York: Mouton de Gruyter.
Prandi, Michele. 2012. Il congiuntivo e i suoi valori: un bilancio. In Remo Bracchi, Michele Prandi and
      Leo Schena (eds.), *Passato, presente e futuro del congiuntivo*, 97–128. Bormio: Centro Studi Storici
      Alta Valtellina.
Renzi, Lorenzo. 2019. Ancora su come cambia la lingua. Qualche nuova indicazione. In Bruno Moretti,
      Alin Kunz, Silvia Natale and Etna Krakenberger (eds.), *Le tendenze dell'italiano contemporaneo
      rivisitate. Atti del LII Congresso Internazionale della Società di Linguistica Italiana (Berna, 6–8
      settembre 2018)*, 13–33. Milano: Officinaventuno.
Rohlfs, Gerhard. 1969. *Grammatica storica della lingua italiana e dei suoi dialetti. Vol. II: Morfologia.*
      Torino: Einaudi.
Santulli, Francesca. 2009. Il congiuntivo italiano: morte o rinascita? *Rivista italiana di linguistica e di
      dialettologia* XI. 151–80.
Schneider, Stefan. 1999. *Il congiuntivo tra modalità e subordinazione: uno studio sull'italiano parlato*.
      Roma: Carocci.
Schulz, Petra. 2003. *Factivity: Its Nature and Acquisition*. Berlin/Boston: Max Niemeyer Verlag.
Serianni, Luca. 2010 [1989]. *Grammatica italiana. Italiano comune e lingua letteraria*. Torino: Utet.
Squartini, Mario. 2008. Lexical vs. grammatical evidentiality in French and Italian. *Linguistics* 46,
      917–947.
Squartini, Mario. 2010. Mood in Italian. In Björn Rothstein and Rolf Theroff (eds.), *Mood in the
      languages of Europe*, 237–250*.* Amsterdam: John Benjamins.
Stewart, Dominic. 2002. Il congiuntivo italiano: un modo della realtà? Uno sguardo al congiuntivo
      nelle grammatiche italiane moderne. In Leo Schena, Michele Prandi and Marco Mazzoleni,
      105–122. Bologna: CLUEB.
Tagliamonte, Sali A. and R. Harald Baayen. 2012. Models, forests and trees of York English: Was/were
      variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178.
Voghera, Miriam. 1993. La grammatica del LIP. In Tullio De Mauro, Federico Mancini and Massimo
      Vedovelli, *Lessico di frequenza dell'italiano parlato*, 86–111. Milano: Etaslibri.
Wandruszka, Ulrich. 1991. Frasi subordinate al congiuntivo. In Lorenzo Renzi and Giampaolo Salvi
      (eds.), *Grande grammatica italiana di consultazione*, II, 415–481. Bologna: Il Mulino.