

# Relativization strategies and sociolinguistic variation in spoken Italian: a typological account

SILVIA BALLARÈ

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

Submitted: 28/10/2024 Revised version: 18/12/2024

Accepted: 20/12/2024 Published: 23/01/2025



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

## Abstract

In this paper I aim at describing and analysing relative clauses in a corpus of spoken Italian. In the first section, I provide an overview of the relativization strategies in Italian, also taking into account non-standard varieties; then, I briefly discuss the sociolinguistic characterization of the sub-standard area of contemporary Italian. In § 2, I introduce the selected corpus and its characteristics, also explaining the methodologies adopted for data extraction and annotation. Then, in § 3, the results of the analysis are presented. The distributions of the different strategies and the outputs of a statistical analysis show the different importance assumed by both linguistic and extralinguistic factors and enable the explanation of the observed variability. Finally, in § 4, some general conclusions are drawn.

**Keywords:** relative clauses; sociolinguistic variation; language variation; spoken Italian.

## 1. Framework

Relative clauses are a widely studied topic in linguistics; also recently, much attention has been devoted to these structures from different perspectives (see e.g. Alexiadou et al. 2000; Kidd 2011; Henderey 2012; Ackerman & Nikolaeva 2013; Cinque 2020). Even in Italian, the topic has been discussed at length in the literature (see below). Relative clauses in Italian can be realized through an array of different strategies:

speakers have multiple options and both simplification and complexification processes come into play.

In this contribution, I discuss the behavior of relativization strategies of all grammatical relations in a small corpus of informal spoken Italian, involving speakers with different social characterizations. Linguistic and extralinguistic factors will be taken into account to discuss and explain the behavior of these structures in spoken data. In addition, the analysis will be conducted by adopting classificatory categories and notions typical of linguistic typology, given that “the patterns of variation and change found in [...] a particular language are in many cases simply instances of patterns of variation and change found across languages” (Croft 2022: 27).

The sociolinguistic analysis is intertwined with the adoption of typological theoretical tools to build a bridge between intralinguistic and interlinguistic variation (see Inglese & Ballarè 2023 *inter al.*). The analysis of structural differences displayed by varieties of the same language in a typological perspective on the one hand shows that non-standard variants are not to be considered mere “accidents” as are well attested in other languages and, on the other, it allows for crosslinguistic comparisons.

In the first part of this section, I present relativization strategies in Italian; in the second, adopting a sociolinguistic perspective, I introduce the Italian sub-standard area.

### 1.1. *Relative clauses in Italian*

In standard Italian, relative clauses can be realized through different strategies (see Serianni 2010 [1989]: 217-240). Nominatives and accusatives in non-restrictive relative clauses can be introduced by ART. + *quale* (‘which’), which must be inflected to display gender/number agreement with the antecedent, as in (1). The same grammatical relations (both in restrictive and non-restrictive relative clauses) can be expressed with the invariable *che*<sup>1</sup> (‘that’), as in (2). The first strategy is more formal,

---

<sup>1</sup> Some authors (see Cristofaro & Giacalone Ramat 2007) consider *che* (‘that’) as part of a morphological paradigm composed of two cells filled with *che* (‘that’) and *cui* (‘which’), due to diachronic reasons. However, *che* (‘that’), unlike *cui* (‘which’) that behave almost exclusively as a relativizing element in the standard variety, can be used with different functions. In fact, it can be employed, for example, to introduce completive clauses and adverbial subordinates, and, in the literature, it is often considered to be a “general subordinator” (also ‘multifunctional *che*’, see below). Furthermore, the two elements are placed in two different stages in the “pronominality cline” (lit. *cline di pronominalità*) proposed by Fiorentino (1999: 164). Because of the high polyfunctionality of *che* (‘that’) and, consequently,

and it is typically attested in highly controlled productions, while the latter is more neutral from a sociolinguistic point of view, as it occurs both in high and low productions.

- (1) *Marco parla a Giulia, la quale dorme*  
 Marco talk:PRS.3SG to Giulia(SG.F) DEF:SG.F REL sleep:PRS.3SG  
 ‘Marco talks to Giulia, who sleeps.’

- (2) *Marco parla a Giulia, che dorme*  
 Marco talk:PRS.3SG to Giulia REL sleep:PRS.3SG  
 ‘Marco talks to Giulia, who sleeps.’

All the other grammatical relations can be realized by the means of a preposition that expresses the function of the antecedent in the subordinate clause followed by the invariable *cui* (‘which’)<sup>2</sup>, as in (3), or by the inflected form of ART. + *quale* (‘which’), as in (4). *Cui*, even if not preceded by any preposition, can be used to express genitive if placed inside the noun phrase between the article and the noun; however, this use is quite rare and attested only in highly formal productions.

In addition, Italian, as many other European languages (see Murelli 2011: 184), has a dedicated form to relativize locative values, i.e. *dove* (‘where’), as shown in (5). Some spatial values, such as the ablative, can be expressed by combining a preposition with *dove*, as in *da dove* ‘from where’. Lastly, most grammars consider standard the employ of *che* (‘that’) to relativize a temporal value, as in (6).

- (3) *La ragione per cui sono in ritardo è*  
 DEF:SG.F reason:SG.F for REL be:PRS.1SG late be:PRS.3SG  
*il maltempo*  
 DEF:SG.M bad.weather:SG.M  
 ‘The reason why I am late is bad weather.’

---

differences in the breadth of functional domains between *che* (‘that’) and *cui* (‘which’), these two elements are treated as independent. By favoring a synchronic approach, in fact the (historical) opposition of these two elements is in the process of being lost or, at the very least, weakened.

<sup>2</sup> Please note that, in this paper, *cui* and *quale* are both translated as *which*. However, as illustrated in this section, if preceded by a preposition, they can be both employed also as *whom*, while *cui* can be used also as *whose*. Due to differences in the relativizing strategies between Italian and English, the translation can be misleading.

- (4) *La ragazza della quale ti ho parlato*  
 DEF:SG.F girl:SG.F of.DEF:SG.F REL DAT.2SG speak:PST.1SG  
*è Anna*  
 be:PRS.3SG Anna  
 ‘The girl I spoke to you about is Anna.’
- (5) *La città dove vivo è Milano*  
 DEF:SG.F city:SG.F REL live:PRS.1SG be:PRS.3SG Milan  
 ‘The city where I live is Milan.’
- (6) *Il giorno che ti ho conosciuto pioveva*  
 DEF:SG.M day:SG.M REL DAT.2SG AUX.1SG know:PST.PTCP rain.PST.3SG  
 ‘The day I met you it was raining.’

The whole paradigm of relativization strategies in standard Italian is summarized in table (1).

	no agreement with head noun	agreement with head noun
NOM. and ACC.	<i>che</i>	ART. + <i>quale</i>
LOC.	PREP. + <i>cui</i> (PREP. +) <i>dove</i>	PREP. + ART. + <i>quale</i>
TEMP.	PREP. + <i>cui</i> <i>che</i>	PREP. + ART. + <i>quale</i>
All other relations	PREP. + <i>cui</i>	PREP. + ART. <i>quale</i>

**Table 1:** Relativization strategies in standard Italian.

The relativization strategies attested in Italian can be categorized through the taxonomy proposed by Comrie & Kuteva (2013a, 2013b) in the World Atlas of Language Structures. In this perspective, standard Italian displays:

- two<sup>3</sup> *relative pronoun strategies* that involve ART. + *quale* and *cui*, in which the element is case marked by a preposition (or by its absence, since Italian does not have a dedicated adposition to express nominatives and accusatives) to

<sup>3</sup> One could also add the case of PREP. + *dove* (‘where’), given that the preposition expresses the locative value.

indicate the role of the antecedent within the subordinate clause; more precisely, ART. + *quale* can be used to relativize nominative and accusatives, while PREPOSITION + *cui* ('which') or ART. + *quale* ('which') can be used to relativize all other grammatical relations.

- two *gap strategies*, realized through *che* ('that') and *dove* ('where'), where there is no overt case-marked reference to the head noun within the subordinate clause. *Che* ('that') is semantically empty (and, in fact, it is used to introduce different kinds of subordinate clauses such as the completive ones), while the meaning of *dove* ('where') is linked with locative values, just as English *where*; however, from a structural point of view, they are invariable and neither of them is case marked (by case or by an adposition).

The array of relativization strategies is much wider when taking into account non-standard varieties of Italian. First, some of the aforementioned elements have broadened their functional domain, and, thus, are used to express more values than in the standard variety. The invariable *che* ('that') is not rarely employed to relativize obliques, as in (7); *dove* ('where') is used with non-spatial antecedent and, sporadically and especially in interactions that involve speakers with low educational achievements, it can be used to relativize nominatives, as in (8) (for a detailed discussion see Ballarè & Inglese 2022).

(7) Alfonzetti 2022: 59 cit. in Cerruti (2017: 65)

*Non c' è nessuno che posso chiedere?*  
 NEG there be:PRS.3SG nobody REL can:PRS.1SG ask:INF  
 'Is there anyone I can ask?'

(8) Bernini 1989: 91

*Nel greco c' è un dativo dove  
 in. DEF:M.SG Greek.M.SG there be:PRS.3SG INDEF:SG.M dative:M REL  
 può presentare una enne finale  
 can.PRS.3SG show:INF INDEF:SG.F n:SG.F final:SG.F*  
 'In Greek there is a dative that can show a final *n*.'

Furthermore, as discussed in detail by Cerruti (2017), in non-standard varieties a wider range of structural possibilities is attested. It is worth noting at least 2 additional constructions.

The first one involves an invariable element - typically *che* ('that'), as in (9), but sporadically also *dove* ('where'), as in (10) - followed by a clitic pronoun which provides

information about the grammatical relation of the relativized element and in some cases agrees with it in terms of gender and number; this is true for datives, that show gender/number agreement, but not for locatives. When a nominative is relativized, given that Italian does not have subject clitics, a tonic pronoun is retained, as in (11). These cases are classified as *resumptive pronoun* by Comrie & Kuteva (2013a, 2013b).

(9) Alfonzetti 2002: 59 cit. in Cerruti (2017: 66)

*I due americani che gli ho aperto*  
 the:PL.M two american:PL.M REL DAT.3PL AUX.1SG open:PST.PTCP  
*l' ombrellone*  
 DEF:SG.M beach.umbrella:SG.M

'The two Americans for whom I opened the beach umbrella.'

(10) KIParla Corpus, PTD012

*Una strada [...] dove ci passa molta più*  
 INDEF:SG.F street:SG.F REL LOC pass.by:PRS.3SG much:SG.F more  
*gente*  
 people:SG.F

'A street where much more people pass by.'

(11) Berretta 1993: 232 cit. in Cerruti (2017: 66)

*c' era [...] Cesarini, che lui all' ultimo*  
 there be:PST.3SG Cesarini REL SUBJ.3SG at.DEF:SG.M last:SG.M  
*minuto faceva sempre goal*  
 minute:SG.M do:PST.3SG always goal

'There was Cesarini, who always scored a goal at the last minute.'

Lastly, there are cases in which there is a *double encoding* (Murelli 2011) of the grammatical relation of the relativized element. More specifically, the construction consists of one inflected element (i.e. PREP. + ART. + *quale* or *cui*) followed by a clitic pronoun that re-expresses the grammatical relation of the antecedent in the subordinate clause, as in (12).

(12) itTenTen20 corpus

*Sembravo un bambino a cui gli era*  
 seem:PST.1SG INDEF:SG.M child:SG.M to REL DAT.3SG AUX.3SG

<i>stato</i>	<i>fatto</i>	<i>il</i>	<i>regalo</i>	<i>che</i>	<i>da</i>	<i>sempre</i>
PST.PTCP	do:PST.PTCP	DEF:SG.M	gift:SG.M	REL	from	always
<i>aveva</i>	<i>desiderato</i>					
AUX.3SG	desire:PST.PTCP					

'I looked like a child who had been given a gift that he had always desired.'

## 1.2. Sociolinguistic variation: the sub-standard area

As it has been shown, Italian displays a complex set of relativization strategies; in fact, different grammatical relations must be realized through different strategies and more strategies can be employed for the same grammatical relation.

Not surprisingly, these strategies display a different sociolinguistic characterization, and this setting lends itself well to numerous studies that, over the years, shed light over its variability and sociolinguistic variation (see Alisova 1965; Alfonzetti 2002; Fiorentino 1999; Cerruti 2016, 2017 inter al.). Suffice to say that in the seminal work authored by Berruto (2012) on sociolinguistic variation in contemporary Italian, relative clauses are emblematically selected as case study to give account for morphosyntactic variation (Berruto 2012: 48ff.). More specifically, the scholar creates a continuum in which he displays all the strategies by crossing two ordered dimensions: the first one is composed of different varieties of Italian- from the higher pole of written standard to the lower of *Italiano popolare* (lit. 'popular Italian', see below)- and the other one gives account of structural characteristics -from the synthetic to the analytic pole.

Regrettably, the persistent lack of freely accessible spoken corpora providing speakers' metadata that has characterized the Italian scenario over the years has led to difficulties in confronting systematically different varieties of Italian and in allowing for further reading of the data, especially from a quantitative perspective. This situation has been changing over the last few years, also thanks to the publication of the KIParla corpus (Mauri et al. 2019) which is a freely accessible resource consisting in spoken data accompanied by a large set of metadata (see § 2), allowing for the analysis of sociolinguistic variation.

In this study, I aim at describing and analyzing how relative clauses are realized in sub-standard productions. More specifically, informal spoken interactions involving speakers with different social characterizations will be taken into account. As it is well known, the informal style is the one in which speakers more easily distance themselves from the standard and, thus, allow us to investigate more in depth the behavior of deviant

strategies; furthermore, the social dimension will be considered because, traditionally, it has been considered highly explicatory to give account for sociolinguistic variation.

The social dimension has been of great relevance in identifying a very important variety within the architecture of contemporary, i.e. the so-called *Italiano popolare* (lit. 'popular Italian'). This variety has been identified in the Seventies (De Mauro 1970; Cortelazzo 1972) and it has been associated with speakers with low educational level that have an Italo-romance dialect<sup>4</sup> as a mother-tongue and employ Italian only in more controlled contexts, where the use of dialect would be strongly stigmatized. The visibility of *Italiano popolare* has greatly diminished in recent decades, and in the literature the scope of the label has been downplayed or the very existence of the variety has been denied (see Lepschy 2002; Renzi 2000, 2012). In support of these, following Berruto (2014: 278-279), two main arguments can be identified. One argues that there are no longer prototypical speakers of *Italiano popolare* and the other that the linguistic features that characterized *Italiano popolare* are to be considered generically sub-standard since they systematically appear in informal productions, regardless of the social characterization of the speakers.

The other main sub-standard variety is the so-called *colloquial Italian*, which is used in everyday, spoken but also written, interactions by speakers of various social characterization, included the ones with higher educational achievements (Berruto 2012 [1987]: 163; Ballarè 2024). From a sociolinguistic perspective, colloquial Italian is maximally relevant because it constitutes, along with *Italiano popolare*, the privileged place where linguistic innovations arise and thus the space in which ongoing variation can be observed.

In this paper, thanks to the analysis of relative constructions, it will be discussed if (and how) speakers with diverse social characterizations behave in different ways in informal spoken productions; this will allow us to discuss if, at least for relativization strategies, the sub-standard part of the architecture of contemporary Italian is homogenous or if there are relevant differences that allow us to distinguish different linguistic behaviors.

---

<sup>4</sup> Italo-romance dialects are different languages from Italian and *not* (geographical) varieties of Italian, in that they all derive from Latin and, thus, they display a structural distance from standard Italian that is similar to the one that can be found, for example, between Spanish and French. For a discussion regarding the structural characteristics of Romance languages, including Italo-romance dialects (such as the dialects of northern Italy -Benincà et al. 2016- and the one of southern Italy - Ledgeyway 2016) see Ledgeyway & Maiden (2016).



## 2. Data and methods

In this section, the ParlaTO corpus is briefly presented together with the choices that were made to identify two subcorpora; then, the methodology adopted in order to extract and code the data is explained.

### 2.1. *ParlaTO* corpus

The ParlaTO corpus (Cerruti & Ballarè 2021) is a module of the larger KIParla corpus (Mauri et al. 2019). It consists of semi-structured interviews collected in the urban area of Turin with speakers balanced by age group (16-29, 30-59, over 60) diversified by social characteristics (gender, educational achievement, occupation). The corpus consists of 48:51 hours of total recordings, 65 interviews and 552.461 tokens. For the purpose of this study, it is important to specify that the interviews, in the vast majority of cases, were conducted by students/researchers who were familiar with the informants (there are, for example, interviews involving relatives and friends) or otherwise in the presence of an intermediary (i.e., a person who knew both the interviewer and the interviewee, and that, by participating in the interaction, cooperated in making the exchange less controlled). In these interactions, speakers were asked for opinions about the city of Turin (about their neighborhood of residence, the change that had occurred over the years, etc.): the topic was selected because it was hypothesized that it might be of interest to the speakers and might engage them in expressing views and opinions. In addition, the exchanges almost always took place in locations selected by the interviewees themselves so that they could be more comfortable. Although the semi-structured interview is a rather codified type of interaction, due to the methodological choices made during the collection phase (see, e.g., Labov 1984: 32-33), overall, these can be considered as rather informal interactions.

In order to observe social variation, two subcorpora were created:

- Subcorpus L (166.540 tokens): semi-structured interviews with speakers with at the most a secondary school license; all available interviews within the corpus were taken, for a total of 12 interviews with 15 informants.
- Subcorpus H (169.376 tokens): semi-structured interviews with speakers with at least a high school diploma; in order to maximize the distance with the social characterization of the speakers of the other subcorpus, informants with a technical/professional school diploma were excluded. Through a randomization

of the selected interviews, a sample size similar to the previous one in terms of tokens was created, for a total of 18 interviews with 22 informants.

The parameter "educational achievement" was selected to divide the speakers into two groups and, consequently, create the two subcorpora exemplifying the social varieties under scrutiny. More specifically, in subcorpus L there are 9 speakers with a primary school license and 6 with a secondary school license; in subcorpus H, on the other hand, there are 7 speakers with a high school diploma, 6 college students and 9 college graduates. Among the available metadata, educational achievement was selected to create socially differentiated groups, as traditionally done in the literature (cfr. Berretta 1988). In fact, a different degree of education often correlates with morphosyntactic variation (see Berruto 1983 *inter al.*). Furthermore, note that *Italiano popolare* is identified *per definitionem* taking into account speakers' educational achievement.

Speakers of the two subcorpora, moreover, are also diversified by age group (and thus employment) and geographical origin, as shown in table (2).

	Subcorpus L	Subcorpus H
Age range		
21-30	1	6
31-40	0	6
41-50	0	2
51-60	0	3
61-70	4	3
71-80	5	1
Over80	5	1
Occupation		
Retailers	0	2
Managers and directors	0	1
Laborers	1	0
Pensioners	14	4
University students	0	6
Geographic origin		
North	8	19
Center	0	1
South and islands	7	2

**Table 2:** The social characterization of the speakers.

Looking at the values shown in the table, it is clear that in subcorpus L there are almost exclusively speakers over 60 years old (and, therefore, pensioners), half of whom were born in northern regions and the other half in southern regions. The picture is quite different in subcorpus H, where there are speakers of different age groups (from 21-30 to over80), who have various jobs and who in the vast majority of cases were born in northern regions. These differences must be linked to the fact that young people are, generally, higher educated and that a massive immigration from southern to northern regions took place in Italy from the 1950s to the 1970s.

The whole corpus has a small dimension and consists of 335.916 tokens; this is due to the fact that there were only 12 interviews with speakers with low educational achievements and, thus, already mentioned, in order to create a balanced sample, I decided to take into account a comparable number of tokens also for the subcorpus H. Furthermore, I was forced to use a rather small amount of data because the analysis of the scrutinized linguistic features required a very laborious and time-consuming manual work of data cleaning, given that the KIParla corpus is not tagged. For instance, to analyze the relative clauses realized through *che* ('that'), it was needed to manually select them among all the 6.072 occurrences of the aforementioned linguistic items in the corpus.

## 2.2. Data extraction and annotation

In order to detect all the relative clauses and given that the KIParla corpus is not morpho-syntactically annotated, all the occurrences of (PREP. +) ART. + *quale/i, cui, che* and *dove* were extracted; this led to a datafile composed of 6.973 occurrences that has been manually cleaned, ruling out:

- a) Cases in which *che* ('that') and *dove* ('dove') were not used as relativizing elements but, for example, as complementizer for completive clauses or as interrogative pronoun/adverb in questions.

This selection was not always straightforward because of cases of the so called *che polivalente* (lit. 'multifunctional *che*', see Fiorentino 2011), that can introduce relative clauses or other subordinates. In order to disambiguate, all the cases in which *che* ('that'), according to the standard rules, could be replaced by another relativizing element were taken into account, as is the case of (13), in which, for example, *in cui* ('in which') could be used in place of *che* ('that').

(13) KIParla, PTD009

*Non è che viviamo in Olanda, che con quattro  
 NEG be:PRS.3SG COMP live:PRS.1PL in Holland REL with four  
 gradi sotto zero prendi la bicicletta  
 degrees below zero take:PRS.2SG DEF:SG.F bicycle:SG.F*

‘It is not like we live in the Netherlands, where you take the bicycle with four degrees below zero.’

b) Cases in which the relative clause was not fully realized, in that the speaker introduced the subordinator (i.e. the relativizing element) but then, the main verb is not produced and, thus, it was not possible to identify univocally the grammatical relation conveyed by the relativized element.

c) Occurrences realized by the interviewer (and not by interviewee).

This process has resulted in a datafile composed of 2.898 sentences that were manually annotated according to the following features.

First, the linguistic element employed was considered, in order to allow the discussion of their sociolinguistic characterization.

a) Relativizing element:

- i. ART. + *quale* (‘which’);
- ii. *cui* (‘which’);
- iii. *che* (‘that’);
- iv. *dove* (‘where’).

The sociolinguistic standardness of the occurrence was also annotated, using as a reference the Italian grammar authored by Serianni (2010 [1989]).

b) Sociolinguistic standardness:

- i. standard;
- ii. sub-standard.

Each occurrence was also tagged according to the strategy employed, adopting the taxonomy presented in § 1.

c) Strategy (strategy):

- i. relative pronoun;
- ii. gap;
- iii. pronoun retention;
- iv. double encoding.

Then, other linguistic features, semantic and syntactic in nature, that have traditionally been considered relevant in explaining variation and relativizing strategies were taken into account.

The grammatical relation that linked the antecedent with the relative clause was annotated, in order to verify with which strategies they were relativized. Nominative and accusative have been merged, given that they exhibit very little variability (see below) and can be relativized with the same strategies, since, as already mentioned, Italian does not have any dedicated preposition to mark nominative and accusative. Dative is expressed by the means of *a* ('to'), while genitive by *di* ('of') both followed by ART. + *quale* or *cui* ('which'). Locatives show a more heterogeneous behavior: several prepositions (followed by ART. + *quale* or *cui* 'which') can be employed, depending on the configuration of the described event, and *dove* 'where' alone can be selected. Furthermore, even though it is not traditionally considered a grammatical relation, we added the temporal value. This value is expressed by the means of a preposition (typically *in* 'in') followed by *quale* or *cui* ('which') or, differently from other non-nominative/accusative grammatical relations, by *che* ('that'). Given this latter structural possibility, we decided to control its behavior separately from other oblique relations.

For the sake of brevity, in the rest of the paper I will refer to the relations from ii. to vi. in d) as *obliques*; however, here *oblique* is to be understood as 'grammatical relations that *can* be relativized by the means of a preposition'. This label, basically, excludes only nominatives and accusatives, given that Italian does not have prepositions that express these grammatical relations.

d) Grammatical relation:

- i. nominative and accusative;
- ii. dative;
- iii. genitive;
- iv. locative;
- v. temporal;
- vi. other.

Then, all the occurrences were coded considering if the relativized element was an argument or an adjunct, in order to verify if the bond with the verb had a relevance in selecting the relativization strategy. As is well known, most of the arguments are nominative, accusative or dative but they can include also locative when a motion or a stative verb is involved.

## e) Argument structure:

- i. argument;
- ii. adjunct.

Furthermore, all oblique relative clauses were annotated according to their semantics, i.e., it was tagged whether they were restrictive or non-restrictive, considering that in the former case they are considered to be more syntactically integrated within the sentence. As is well known, restrictive relative clauses allow for the identification of a referent among a set of possible referents, while non-restrictive relative clauses provide additional information about a referent.

## f) Semantics:

- i. restrictive;
- ii. non-restrictive.

In order to distinguish the two categories, sentence negation was adopted as main criterion. As discussed by Cristofaro (2005: 195-196), negating a sentence containing a restrictive relative does not negate the content of the relative itself, as in (14a), while more interpretations are allowed when negating a sentence containing a non-restrictive relative clause, as in (14b).

(14) adapted from Cristofaro (2005: 195)

a. *The man [who is sitting in that office] is a psychologist.*

→ It is not true that he is a psychologist.

b. *They went to a number of Bach concerts, [for which they had booked tickets several months in advance].*

→ It is not true that they went to a number of Bach concerts; it is not true that they had booked tickets several months in advance; it is not true that they went to a number of Bach concerts, neither that they had booked tickets for them several months in advance.

Finally, other two linguistic parameters were annotated, in order to verify if they could play a role in the selection of the relativization strategy. First, I considered the target prepositions to verify if their diverse frequencies had consequences on the employed strategy. Then, I took into account the definiteness of the antecedent to verify whether a greater degree of accessibility favors the selection of more explicit syntactic strategies.

- g) Preposition:
  - i. *a*, ‘to’;
  - ii. *con*, ‘with’;
  - iii. *da*, ‘from’;
  - iv. *di*, ‘of’;
  - v. *fra/tra*, ‘between’ or ‘among’;
  - vi. *in*, ‘in’;
  - vii. *per*, ‘for’;
  - viii. *su*, ‘on’;
  - ix. *riguardo (a)*, ‘about’.
- h) Definiteness of the antecedent:
  - i. definite;
  - ii. indefinite.

The main objective will be to discuss whether speakers with different social characterization use structurally different strategies for relativization. It will be considered whether and how different linguistic factors have relevance in the selection of different relativization strategies.

### 3. Discussion

After a brief overview over the frequencies of relative clauses in the two sub-corpora (H and L), the behavior of nominative/accusative and obliques will be discussed.

In table (3) are reported the absolute values of relative clauses in the two sub-corpora, taking into account their grammatical relation. Here and in the following tables, percentage values are displayed in brackets.

	Nom and Acc	Obliques	Tot.
H	1.445 (85,91%)	237 (14,09%)	1.682 (100%)
L	1.006 (82,73%)	210 (17,27%)	1.216 (100%)
			2.898

Table 3: Distribution: grammatical relations.

The first thing that can be noted is that relative clauses are more frequent in the productions of highly educated speakers. This is shown by the absolute values (1.682

vs. 1.216) and it is confirmed by the relative frequencies<sup>5</sup>, which are 9,93 in H and 7,30 in L.

If we consider the distribution of the relative clauses in the two sub-corpora between the 2 types of grammatical relations, we note that the values are similar, even if some differences can be high-lightened. The vast majority of occurrences involve nominatives and accusatives, while all other cases are relativized more sporadically. However, speakers with higher educational achievements, proportionally, relativize nominatives and accusatives more frequently than the others (85,9% vs. 82,7%); and, specularly, speakers of the L corpus, proportionally, relativize obliques more often (17,3% vs. 14,1%). An analogous result has been observed comparing formal and informal spoken productions of Italian and in other languages (see Ballarè & Larrivée 2021); one could hypothesize that in lower productions speakers prefer to employ strategies different from relative clauses to modify a nominal head (such as the repetition of the nominal head itself) but further studies are needed.

Globally, the distribution is statistically significant at  $p < 0,05$  (Fisher exact test statistic value is 0,0218).

### 3.1. Nominative and accusative

In this section the focus is on the relativization of nominative and accusative; in table (4) there are displayed the strategies selected in the two subcorpora. No cases of relative pronoun (i.e. ART. + *quale* ‘which’ and inflected variants) are attested and double encoding is not one of the options given that in Italian there is no case marking for nominative and accusative.

	Gap	Resumptive pr.	Tot.
H	1.438 (99,52%)	7 (0,48%)	1.445 (100%)
L	997 (99,11%)	9 (0,89%)	1.006 (100%)
			2.451

**Table 4:** Distribution: strategies (nominative/accusative).

<sup>5</sup> (number of occurrences / number of tokens of the sub-corpus)\*1000.



The gap strategy is the one selected almost categorically. In the productions of highly educated speakers, it involves *che* ('that') in all the cases but 2, in which one speaker relativizes two nominatives selecting *dove* ('where'), as exemplified in (15). In L, an analogous situation is observed: *che* ('that') is selected in 994 cases over 997 and there are 3 occurrences of *dove* ('where') to relativize a nominative, as in (16). It is worth noting that in all the 5 cases in which *dove* ('where') is involved, the nominal antecedent is a location -as in (16)- or it is a derived form of a spatial noun, as in (15) where *meridionale* ('southerner') derives from *meridione* ('south').

(15) KIParla, PTB019

<i>con</i>	<i>il</i>	<i>meridionale</i>	<i>dove</i>	<i>abitava</i>	<i>in</i>	<i>via</i>	<i>Montenero</i>
with	DEF:SG.M	southerner:SG.M	REL	live:PST.3SG	in	street	Montenero

'With the southerner who lived in Montenero street.'

(16) KIParla, PTA005

<i>poi</i>	<i>hai</i>	<i>il</i>	<i>bar</i>	<i>del</i>	<i>cinese</i>
then	have:PRS.2SG	DEF:SG.M	bar:SG.M	of. DEF:SG.M	chinese:SG.M
<i>dove</i>	<i>però</i>	<i>ha</i>	<i>una</i>	<i>sua</i>	<i>clientela</i>
REL	but	have:PRS.3SG	INDEF:SG.F	GEN.3.SG.F	clientele:SG.F

'Then you have the Chinese's bar, that has its clientele.'

If we consider the data, we can see that speakers with different educational achievements behave in a homogeneous way in informal productions and there are no significant differences<sup>6</sup>. This is true for the adopted strategies and selected linguistic items. That is to say that relativization strategies of nominative and accusative in informal spoken Italian are uniform regardless of the social characterization of the speakers. In fact, ART. + *quale* ('which') is completely absent and the employ of *che* ('that') is almost categorical. There are globally only 21 sub-standard occurrences out of 2.451, consisting of the employ of *dove* ('where') to relativize subjects (5 occurrences) and the co-occurrence of a pronoun with *che* ('that') (16 occurrences); these last occurrences always involved the

<sup>6</sup> The Fisher exact test statistic value is 0,3076 and the result is thus not significant at  $p < 0,05$ .

relativization of an accusative and the employ of a clitic pronoun, except in one case, reported in (17).

(17) KIParla, PTB002

<i>Mi</i>	<i>son</i>	<i>fermato</i>	<i>tante di quelle volte</i>	<i>da</i>
REFL.1SG	AUX.1SG	stop:PST.PTCP	many.times	to
<i>questo</i>	<i>mio</i>	<i>amico</i>	<i>che lui</i>	<i>tante volte</i>
DEM.SG.M	POSS.1SG	friend:SG.M	REL SUBJ.3SG	many times
<i>usciva</i>	<i>con</i>	<i>la</i> <sup>7</sup>		
go.out:PST.3SG	with	DEF:SG.F		

'I stopped many times at this friend of mine that used to go out with (her).'

### 3.2. Obliques

#### 3.2.1 Distributions

As mentioned, the relativization of the obliques is where greater variability is expected. First, let us consider the distribution of non-standard realizations in the two sub-corpora presented in table (5).

	Standard	Sub-standard	Tot.
H	198 (83,54%)	39 (16,45%)	237 (100%)
L	115 (54,76%)	95 (45,24%)	210 (100%)
			447

Table 5: Distribution: standardness (obliques).

It is possible to observe how speakers in this case behave in diverse ways: while in H sub-standard realizations constitute only 16,46% of the occurrences, in L they are nearly half of the sample (45,24%). The distribution is statistically significant at  $p < 0,01$  (Fisher exact test statistic value is  $< 0,000001$ ).

<sup>7</sup> Unfortunately, the only example in which this strategy appears is a case of unconcluded utterance. Thus, it is not possible to complete the prepositional phrase. The presence of the definite feminine article (*la*) may lead us to think that the speaker wanted to mention a female person.

It is important to say that the non-standardness of the occurrences may be linked to the relativizing element or, more rarely, the selected preposition. In the rest of the section, the issue will be addressed more in depth.

Let us consider the structural strategies employed in the two sub-corpora in the relativization of the obliques reported in table (6).

	Rel. pron.	Gap	Res. pron.	Double enc.	Tot.
H	98 (41,35%)	125 (52,74%)	10 (4,22%)	4 (1,69%)	237 (100%)
L	19 (9,05%)	181 (86,19%)	10 (4,76%)	0 (0%)	210 (100%)
					447

**Table 6:** Distribution: strategies (obliques).

Overall, looking at the distribution of different relativization strategies in the two subcorpora, we can see macroscopic differences. In both cases, the gap strategy is the most frequently used: however, while in subcorpus H it is employed in just over half of the cases (52,74%), in subcorpus L it exceeds 86%. The second most frequently used strategy is the one involving a relative pronoun; again, however, the frequency values are very different: in H it exceeds 40% while in L it does not reach 10%. The remaining structures, i.e. resumptive pronoun and double encoding, are much rarer; interestingly, the double encoding (i.e. the double expression of the grammatical relation) is only attested in the productions of speakers with higher educational achievements (see Berretta 1993: 232). One example of resumptive pronoun strategy employed by a speaker with lower educational achievements is provided in (18).

(18) KIParla, PTB009

*Tuo papà e l' Elsa che la nonna*  
 POSS.2SG father.SG.M and DEF.SG.F Elsa REL DEF.SG.F grandmother.SG.F  
*Lidia gli insegnava la matematica*  
 Lidia DAT.3PL teach:PST.3SG DEF.SG.F mathematics.SG.F  
 'Elsa and your father, to whom grandmother Lidia taught mathematics.'

Speakers with low educational achievements prefer the only structure that does not involve case marking (i.e. gap); strategies involving a preposition or a clitic pronoun, overall, do not reach 14% of occurrences. Higher-educated speakers, on the other

hand, have more diverse behavior: although the gap strategy is the one employed most frequently, the others (i.e. relative pronoun, resumptive pronoun and double encoding) exceed 47%. A more detailed analysis of these differences will be addressed in the next section.

Before discussing the differences in terms of overt case-marking of the relativized item, it may be useful to look at the linguistic items selected by speakers with different educational achievements, that are reported in table 7. The topic, of course, ties in with the previous one since the different structures cannot be expressed by all the relativizing elements. In fact, one can have a case of relative pronoun or double encoding only with ART. + *quale* ('which') and *cui* ('which') -both preceded by a preposition-, while one can have gap and resumptive pronoun only with *che* ('that') and *dove* ('where')<sup>8</sup>.

	ART. + <i>quale</i>	<i>Cui</i>	<i>Dove</i>	<i>Che</i>	Tot.
H	2 (0,84%)	100 (42,19%)	106 (44,73%)	29 (12,24%)	237 (100%)
L	1 (0,48%)	18 (8,57%)	84 (40,00%)	107 (50,95%)	210 (100%)
					447

Table 7: Distribution: relativizing element (obliques).

In line with what was observed for nominative and accusative, in this case, occurrences of the ART. + *quale* ('which') are rare (3 in total) in both H and L. However, we can note at least two major differences. The first is that *cui* ('which') is much more frequent in H (42,19% vs. 8,57%), and is the form selected for the expression of relative pronouns and double encodings. The second is that speakers in H prefer *dove* ('where') over *che* ('that') for the realization of gap strategy (44,73% vs. 12,24%); specularly, speakers in H use *che* ('that') more frequently than *dove* ('where'); *che* ('that') alone, in fact, is employed to relativize more than half of the obliques in the sub-corpus.

### 3.2.2 Explaining variation

Because of what was observed in the previous paragraph, it is of interest to discuss the differences in speakers' behavior by distinguishing between relativization

<sup>8</sup> Please note that no occurrences of PREP. + *dove* ('where') are attested in the corpus.

strategies that exhibit case marking (i.e. relative pronoun, pronoun retention and double encoding) and those that do not (gap).

In the literature, the topic has been approached in terms of *explicitness*, that is, how explicitly the strategy encodes the role of the antecedent (Comrie 1989: 163). Explicitness is described as gradual but, in our case, also because of the rather small dimension of the dataset, the parameter is treated as binary. As diverse as they are in terms of both structure and standardness, what is of interest here is the need to divide the strategies between those that involve an element, be it a clitic pronoun or be it a preposition, to make explicit the grammatical relation between the antecedent and subordinate clause and the others.

In this section, I discuss the results of a statistical analysis conducted by associating two values, i.e. case marked vs. non-case marked, to the scrutinized variable. The factors considered are (see § 2.2):

- 1) educational achievements of the speaker;
- 2) grammatical relation;
- 3) argument structure;
- 4) preposition;
- 5) definiteness of the antecedent.

The data will be analyzed adopting a conditional inference tree and a random forest (Tagliamonte & Baayen 2012; Levshina 2015), which are indicated when the dataset is unbalanced and rather small. A conditional inference tree is a decision tree used to model relationships between a target variable and more predictor variables. They use statistical tests to decide where to split the data in homogeneous sub-sets: the process involves selecting a predictor variable that has the strongest association with the target, then partitioning the data based on thresholds in that predictor. A random forest builds multiple conditional inference trees and combines their outputs to improve predictive accuracy. The result is a ranking of the selected parameters according to their importance.

The conditional inference tree is shown in figure (1). The C index is 0,83 and thus the model offers an excellent discrimination (Hosmer & Lemeshow 2000: 162).

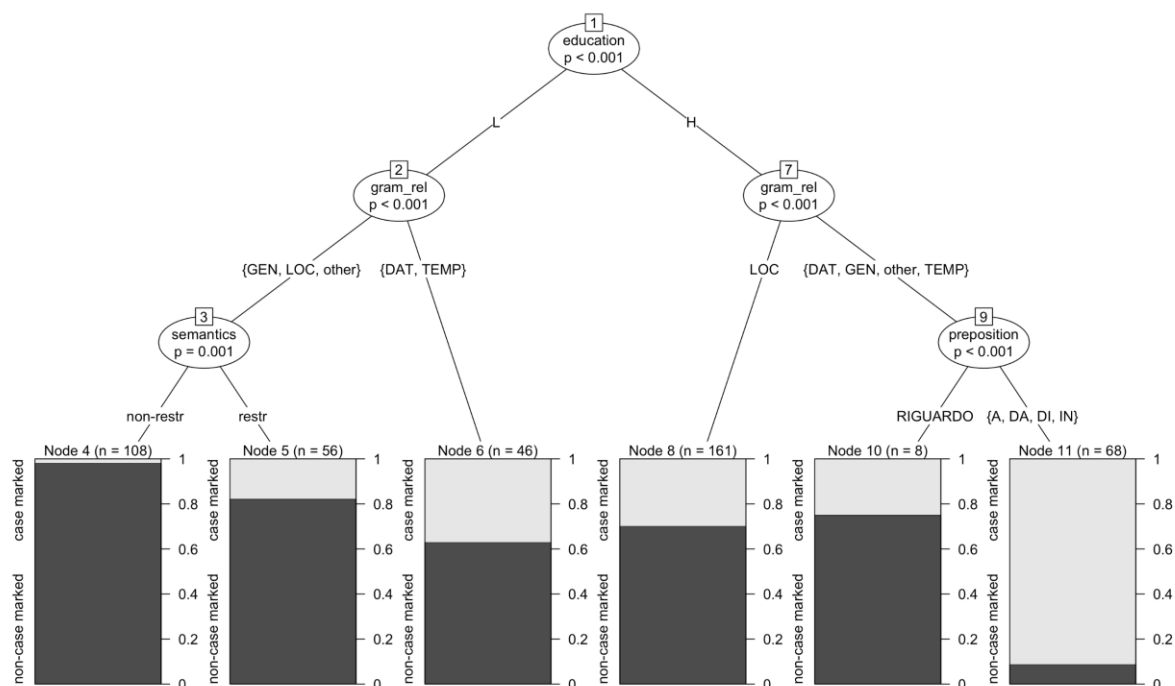


Figure 1: Conditional inference tree.

The first parameter that is of interest in creating homogeneous sub-part of the dataset is the social one, i.e. the educational achievement of the speakers (node 1). The behavior of speakers with lower educational achievements is represented on the left; the one of speakers with higher educational achievements, on the right. Overall, as already noted above, constructions with overt case marking are more frequent in the speech of subcorpus H, that is to say that speakers with higher educational achievements prefer using more complex structures in order to overtly express the grammatical relation relativized by the subordinate clause.

The second parameter relevant in both L and H is the grammatical relation of the relativized element. However, the partitioning of the data is done by grouping the values of the variable differently. In node 2, the division occurs between genitive, locative and *other* vs. dative and temporal. The first group is most often realized by the gap strategy. The locative in Italian has the dedicated form *dove* ('where') and this may be why the gap strategy is preferred. In *other* are placed relations that are rarer and sometimes realized through prepositions less frequent than the others and/or improper (see below). Surprising, however, is the placement of the genitive since it has no dedicated relativizing element, but it is still more frequently realized without the overt case marking. In the second group, i.e. dative and temporal, the gap strategy is still the most frequent but, proportionally, the percentage of case-marked structures

is higher. That is to say that these two relations, compared with the others, are more frequently expressed by non-gap strategy. This is explainable for dative, while it is less explicable for temporals, which, at least according to some grammars, in Italian can be relativized using *che* ('that').

At this point, only in the right portion of the tree and only for the grammatical relations of genitive, locative and *other*, the semantics of the relative (node 3) gain relevance. Not surprisingly, in non-restrictive relatives, thus less syntactically integrated in the sentence, non-case marked strategies (node 4) are more frequent than in restrictive ones (node 5).

Let us now consider the behavior of speakers in subcorpus H, where two linguistic factors come into play. The first, as already mentioned, is the grammatical relation (node 7). If the clause relativizes a locative relation, then it is expressed by a gap strategy in most cases. The difference in behavior between this function and the others is easily explained by the aforementioned presence of the dedicated form (node 8). All other relations are more frequently realized with a case marked strategy even if another factor acquires relevance: the target preposition of the subordinate clause (node 9). In fact, speakers do not use a case marking more frequently when an improper preposition, i.e. *riguardo* ('about'), is employed (node 10). Even though the number of occurrences is rather low (8), two things are worth saying. The first is that *riguardo* ('about') is not polyfunctional and much less frequent in the corpus than the other prepositions and therefore probably less easily retrievable by the speaker. In table (8) I show the normalized frequencies<sup>9</sup> of the different prepositions within the ParlaTO corpus.

Preposition	Normalized frequency
<i>A</i> 'to'	1,4%
<i>Da</i> 'from'	0,5%
<i>Di</i> 'of'	1,6%
<i>Con</i> 'with'	0,4%
<i>In</i> 'in'	1,1%
<i>Per</i> 'for'	0,7%
<i>Su</i> 'on'	0,1%
<i>Tra/fra</i> 'between' or 'among'	0,1%
<i>Riguardo (a)</i> 'about'	0,0027%

**Table 8:** Prepositions' frequencies.

<sup>9</sup>Percentage values are reported to the first decimal place, except for *riguardo* ('about') since its value is very low.

The second thing that can be worth mentioning is that the aboutness value is close to the one of the subject from an informative point of view and, thus, this could be one of the reason why it triggers the selection of a gap strategy, which is typically used to relativize subjects; furthermore, it has been observed that in Italian this values is often realized by the means of *dove* ('where') - see Ballarè & Inglese (2022).

If another preposition is involved, speakers in H select almost categorically a case marked strategy (node 11).

We can now consider the importance of factors in explaining the selection of the relativization strategy in the whole dataset. Figure 2 shows the random forest ranking; its C index is 8.6 (excellent discrimination, Hosmer & Lemeshow 2000: 162) and below are the numerical values obtained from the analysis:

- gram\_rel: 0,058;
- education: 0,046;
- semantics: 0,004;
- definiteness: 0,003;
- preposition: 0,001;
- argument structure: 0,001.

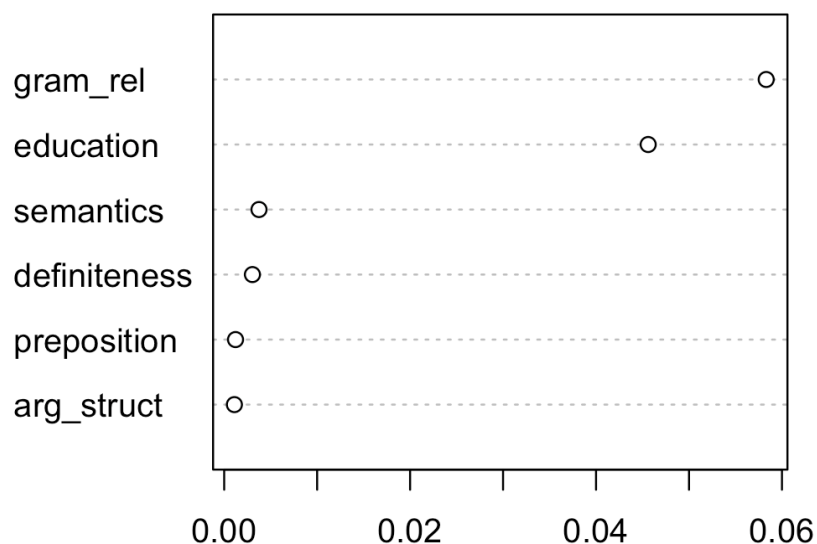


Figure 2: Random forest.

The first two parameters (i.e., grammatical relation and educational achievements) have importance in selecting the relativization strategy; the other four (i.e., semantics, antecedent definiteness, target preposition and argument structure) do not. As can be seen, the random forest indicates the importance of the parameter but not the



direction in which it acts; therefore, below I discuss the distributions associated with the prominent values.

In the first instance, I take into account the values associated with the grammatical relation, shown in table 9. Percentages are calculated based on total column value. The chi-square statistic is 62.087; the *p*-value is < 0,00001, and thus the result is significant at *p* < 0,01.

	<b>dat</b>	<b>gen</b>	<b>temp</b>	<b>other</b>	<b>loc</b>
Case marked	13 (72,22%)	13 (61,90%)	35 (59,32%)	25 (35,21%)	55 (19,78%)
Non-case marked	5 (27,78%)	8 (38,10%)	24 (40,68%)	46 (64,79%)	223 (80,22%)
	18 (100%)	21 (100%)	59 (100%)	71 (100%)	278(100%)

**Table 9:** Strategy and grammatical relation.

Grammatical relations are entered from left to right by those most frequently having overt case marking and decreasing.

Before moving to the analysis, it is important to note that the absolute values of the first two columns are rather low and thus any generalization requires due caution.

Dative and genitive are expressed in the vast majority of cases through the use of a case mark. The temporal and locative values, i.e. the only ones that, according to grammars, admit both strategies, show different behaviors. Temporal values are expressed by employing both available strategies (with a preference for overt case marking), while locative ones are realized more frequently (80,22% of cases) through a non-case marked strategy.

Even if grammatical relations and not semantic roles were tagged, the results can be discussed in relation to a well-known typological generalization. Comrie, in fact, states that “the more difficult a position is to relativize, the more explicit indication is given of what position is being relativized, to facilitate recovery of this information” (Comrie 1989: 163). With “difficulty of relativization”, Comrie is referring to the well-known accessibility hierarchy (Keenan & Comrie 1977: 66), which “is basically the degree of salience of the participant in the relative clause event” (Croft 2022: 604); the further to the left the position, the easier the relativization.

SUBJECT > DIRECT OBJ. > INDIRECT OBJ. > OBLIQUE > GENITIVE > OBJ. OF COMPARISON

Following Comrie (1989: 163), this means that “for instance, where the choice is between a pronoun-retention and a gap relative clause, it is nearly always the case that the pronoun-retention type is used lower down the accessibility hierarchy [...], while the gap strategy is used higher up”.

Considering the Italian data, the generalization remains valid for subject and direct object (approximated to nominative and accusative, with some exceptions) since, as discussed, they are relativized almost categorically with the least explicit structure, namely *che* ('that'). However, the position of the genitive is problematic, since it precedes all the obliques (locative, temporals and *other*) and the indirect object (approximable, again, not without exceptions, to the dative).

The order found, however, albeit with some differences, is reminiscent of another typological hierarchy, namely the inflectional case hierarchy proposed by Blake (2001), reported below.

NOM    ACC/ERG    GEN    DAT    LOC    ABL/INST    *others*

As is known, the hierarchy is to be interpreted as follows: “if a language has a case listed on the hierarchy, it will usually have at least one case from each position to the left” (Blake 2001: 156). Between the scrutinized relativization strategies and the case hierarchy, there are some substantial differences and limits. First, Blake considers only morphological case systems while, in our case, I am referring to a language that, in most cases, expresses grammatical relations by prepositions. Moreover, inflectional case, typically, act on a word level while I am considering strategies that are at work on a syntactic one. The hierarchy also has some limits, since there are several exceptions to it, also given by cases of syncretism (see Baerman et al. 2005 inter al.).

However, when comparing the inflectional case hierarchy and the result of our analysis (disregarding nominative and accusative), it is possible to note the presence first of genitive and dative in a rather high position on the scale, and then the other values. As already mentioned, locatives and temporals in Italian exhibit a different structural behavior compared to other obliques. Although with some differences, what is interesting to note here is that, when possible, speakers proportionally more frequently employ an overt case marking relativization strategy in an order that reminds the one according to which the languages of the world have a dedicated case mark.

The explanation, thus, is not to be found in the accessibility of the participant but more on a syntactic level. The first two grammatical relations, in fact, have a deeper

bond to the main clause with respect to the others: typically, in fact genitive modifies a nominal element, while the dative is a verb argument. The other relations, on the other hand, are, at least prototypically, less bonded and behave more often as adjuncts of the main clause. Syntax, however, also given the non-importance of the argument/adjunct parameter does not tell the whole story.

In fact, it is worth noting that, on a semantic level, dative and genitive, differently from the other grammatical relations, are often linked with animacy<sup>10</sup>. One could argue that, given the higher saliency of animate referent and in order to avoid ambiguity, speakers make explicit the grammatical relation between the antecedent and the relative clause, by selecting a more structurally complex strategy.

The second parameter that shows importance is educational achievements. The distribution is given in table (10); the Fisher exact test statistic value is  $< 0,00001$  and thus the result is significant at  $p < 0,01$ .

	H	L
Case marked	112 (47,26%)	29 (13,81%)
Non-case marked	125 (52,74%)	181 (86,19%)
	237 (100%)	210 (100%)

**Table 10:** Strategy and educational achievement.

As discussed in the previous section, speakers with lower educational achievements prefer to adopt the less complex strategy (86,19%), while those with higher educational achievements show values that hover around 50% for both types of strategies at their disposal; thus, proportionally, they use the more complex strategies more often and explicate the syntactic relation between the antecedent and subordinate clause.

Even though the statistical model considers all other parameters to be not important in explaining how speakers select different relativization strategies, it may be useful to note that the distribution of two of these, i.e., semantics and argument structure (cfr. Fiorentino 1999: 167), turns out to be statistically significant (with  $p < 0,01$  and Fisher exact test value of 0,0007 and 0,0036, respectively). The apparent contradiction is actually easily resolved by clarifying the differences statistical

<sup>10</sup> The relevance of animacy in explaining cases of *pronoun retention* has been noted also by Berretta (1993: 232-233) and Fiorentino (1999: 104).

significance and model importance. In fact, statistical significance measures whether a variable's distribution differs from random chance, often in isolation, while random forest importance assesses a variable's contribution to prediction accuracy within the context of all other variables.

In order to understand more clearly the behavior of the relativizing structures, the distributions of these last two parameters are shown in Tables 11 and 12.

	<b>restr</b>	<b>non-restr</b>
Case marked	75 (40,54%)	66 (25,19%)
Non-case marked	110 (59,46%)	196 (74,81%)
	185 (100%)	262 (100%)

**Table 11:** Strategy and semantics.

	<b>arg</b>	<b>adj</b>
Case marked	49 (42,61%)	92 (27,71%)
Non-case marked	66 (57,39%)	240 (72,29%)
	115 (100%)	332 (100%)

**Table 12:** Strategy and argument structure.

The distributions show quite clearly that when the semantic (i.e. “restrictive”) or syntactic (i.e. “argument”) connection between the relative subordinate and the main sentence is stronger, speakers more frequently select case-marked strategies (40,54% vs. 25,19% and 42,61% vs. 27,71%, respectively).

#### 4. Conclusive remarks

The results of the conducted analysis shed light on relativization strategies in the sub-standard area of spoken Italian.

In general, relative clauses are more frequently realized by speakers with higher educational achievements.

For the relativization of nominative and accusative, the behavior in the sub-standard area is uniform: all speakers, regardless of their social characterization, employ *che* (‘that’). It is important to emphasize that, from a functional point of view, selecting *che* (‘that’) or ART. + *quale* (‘which’) has no substantial consequences given

that in Italian, in this domain, there are no case markers for nominative and accusative. The difference between these two linguistic elements is sociolinguistic in nature: in more controlled contexts, the use of ART. + *quale* ('which') remains frequent, perhaps precisely to mark the formality of the production. There are few occurrences in which a clitic pronoun is also involved, and it is interesting to note that, in terms of frequency, speakers in H and L behave in an analogous way by producing a similar number of sub-standard occurrences, regardless of their educational achievements.

Profound differences have been observed in the relativization of obliques, depending on the social characteristics of the speakers.

Not surprisingly, speakers with lower educational achievements produce more sub-standards relatives; what is of interest, however, is that while they produce them by simplifying the structure and over-extending the gap strategy, speakers with higher educational achievements realize sub-standard occurrences complexifying the structure and employing the double-encoding strategy.

We also notice differences in the selection of relativizing elements: speakers in H frequently use *cui* ('which') and prefer *dove* ('where') over *che* ('that'); speakers in L, on the other hand, use *che* ('that') significantly more frequently. This, of course, ties in with the relativized relations. Statistical analysis showed that there is a significant relation between educational achievements and the adoption of case marked or non-case marked relativization strategies, since highly educated speakers prefer to overtly mark the grammatical relation. Furthermore, the most important factor in selecting a strategy type is the grammatical relation. Genitive and dative, that are syntactically bonded to the main clause and that are the only oblique relations that can involve an animate referent, are the ones more frequently expressed by a case marked strategy.

From a sociolinguistic perspective, we can say that the homogeneity detected for nominative and accusative is *not* found in the obliques because speakers behave significantly differently. That is to say that, at least in our data and at least for the relativization of the obliques, speakers with lower educational achievement select different strategies compared to others. Speakers in H, on the other hand, show greater variability and have more relativization strategies at their disposal.

Studying thoroughly data of a single language from a sociolinguistic perspective allows for an accurate analysis, that also considers specific features of the scrutinized language itself; however, the study shows also that the adoption of typological categories allow us to go beyond them and tie the results to the bigger picture.

## Abbreviations

1 = 1 <sup>st</sup> person	ERG = ergative	PL = plural
2 = 2 <sup>nd</sup> person	F = feminine	POSS = possessive
3 = 3 <sup>rd</sup> person	GEN = genitive	PRS = present
ABL = ablative	INDEF = indefinite	PST = past
ACC = accusative	INF = infinitive	PTCP = participle
ART = article	INS = instrumental	REFL = reflexive
AUX = auxiliary	LOC = locative	REL = relative
COMP = complementizer	M = masculine	SG = singular
DAT = dative	NEG = negation	SUBJ = subject
DEF = definite	NOM = nominative	
DEM = demonstrative	OBJ = object	

## References

- Ackerman, Farrell & Irina Nikolaeva. 2013. *Descriptive typology and linguistic theory: a study in the morphosyntax of relative clauses*. Stanford: CSLI.
- Alexiadou, Artemis, Paul Law, André Meinunger & Chris Wilder (eds.). 2000. *The syntax of relative clauses*. Philadelphia: John Benjamins.
- Alfonzetti, Giovanna. 2022. *La relativa non standard. Italiano popolare o italiano parlato?*. Palermo: Centro di studi filologici e linguistici siciliani.
- Alisova, Tatiana. 1965. Relative limitative e relative esplicative nell'italiano popolare. *Studi di filologia italiana XXIII*. 299-333.
- Baerman, Matthew, Dunstan Brown & Greville G. Corbett. 2005. *The Syntax-Morphology Interface. A Study of Syncretism*. Cambridge: Cambridge University Press.
- Ballarè, Silvia. 2024. L'italiano colloquiale. In Silvia Ballarè, Ilaria Fiorentini & Emanuele Miola (eds.), *Le varietà dell'italiano contemporaneo*, 81-98. Roma: Carocci.
- Ballarè, Silvia & Pierre Larrivée. 2021. Register impacts syntax: scaling the accessibility hierarchy of relatives. *Italian journal of linguistics* 33. 3-22.
- Ballarè, Silvia & Guglielmo Inglese. 2022. The development of locative relative markers from typology to sociolinguistics (and back). *Studies in language* 46. 220-257.
- Benincà, Paola, Mair Parry & Diego Pescarini. 2016. The dialects of northern Italy. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford Guide to the Romance languages*, 185-205. Oxford: Oxford University Press.
- Bernini, Giuliano. 1989. Tipologia delle frasi relative italiane e romanze. In Fabio Foresti, Elena Rizzi & Paola Benedini (eds.), *L'italiano tra le lingue romanze. Atti del XX Congresso*

- Internazionale della Società di Linguistica Italiana (Bologna, 25-27 settembre 1986)*, 85-98. Roma: Bulzoni.
- Berretta, Monica. 1988. Varietätenlinguistik des Italienischen / Linguistica delle varietà. In Günter Holtus, Michael Metzeltin & Christian Schmitt (eds.), *Lexicon der Romanistischen Linguistik. Italienisch, Korsisch, Sardisch*, 762-774. Berlin/New York: Mouton de Gruyter.
- Berretta, Monica. 1993. Morfologia. In Alberto A. Sobrero (ed.), *Introduzione all'italiano contemporaneo I*, 193-245. Roma/Bari: Laterza.
- Berruto, Gaetano. 1983. Italiano popolare e semplificazione linguistica. *Vox Romanica* 42. 38-79.
- Berruto, Gaetano. 2012. *Sociolinguistica dell'italiano contemporaneo (seconda edizione)*. Roma: Carocci.
- Berruto, Gaetano. 2014. Esiste ancora l'italiano popolare? Una rivisitazione. In Paul Danler & Christine Konecny (eds.), *Dall'architettura della lingua italiana all'architettura linguistica dell'Italia. Saggi in omaggio a Heidi Siller-Runggaldier*, 277-190. Frankfurt am Main: Lang.
- Blake, Barry J. 2001. *Case. Second edition*. Cambridge: Cambridge University Press.
- Cerruti, Massimo. 2016. Costruzioni relative in italiano popolare. In Federica Guerini (ed.), *Italiano e dialetto bresciano in racconti di partigiani*, 77-116. Roma: Aracne.
- Cerruti, Massimo. 2017. Changes from below, changes from above. Relative constructions in contemporary Italian. In Massimo Cerruti, Claudia Crocco & Stefania Marzo (eds.), *Towards a new standard. Theoretical and empirical studies on the restandardization of Italian*, 32-61. Berlin/New York: Mouton de Gruyter.
- Cerruti, Massimo & Silvia Ballarè. 2021. ParlaTO: corpus del parlato di Torino. *Bollettino dell'Atlante linguistico Italiano* 44. 13-38.
- Cinque, Guglielmo. 2020. *The syntax of relative clauses: a unified analysis*. Cambridge: Cambridge University Press.
- Comrie, Bernard. 1989. *Language universals and linguistic typology. Second edition*. Chicago: University of Chicago Press.
- Comrie, Bernard & Tania Kuteva. 2013a. Relativization on Subjects. In Matthew Dryer & Martin Haspelmath (eds.), *World atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info/chapter/122>. (Accessed 2024-10-25).

- Comrie, Bernard & Tania Kuteva. 2013b. Relativization on Obliques. In Matthew Dryer & Martin Haspelmath (eds.), *World atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info/chapter/123>. (Accessed 2024-10-25).
- Cortelazzo, Manlio. 1972. *Avviamento critica allo studio della dialettologia italiana, III, Lineamenti di italiano popolare*. Pisa: Pacini.
- Cristofaro, Sonia. 2005. *Subordination*. Oxford: Oxford academic.
- Cristofaro, Sonia & Anna Giacalone Ramat. 2007. Relativization strategies in the languages of Europe. In Paolo Ramat & Elisa Roma (eds), *Europe and the Mediterranean as linguistic areas: Convergences from a historical and typological perspective*, 63–93. Amsterdam/Philadelphia: John Benjamins.
- Croft, William. 2022. *Morphosyntax. Constructions of the World's Languages*. Cambridge: Cambridge University Press.
- De Mauro, Tullio. 1970. Per lo studio dell'italiano popolare unitario. In Annabella Rossi (ed.), *Lettere da una tarantata*, 43-57. Bari: De Donato.
- Fiorentino, Giuliana. 1999. *Relativa debole. Sintassi, uso, storia in italiano*. Milano: Franco Angeli.
- Henderey, Rachel. 2012. *Relative clauses in time and space: a case study in the methods of diachronic typology*. Amsterdam/Philadelphia: John Benjamins.
- Hosmer, David W. & Stanley Lemeshow. 2000. *Applied logistic regression*. New York: Wiley.
- Inglese, Guglielmo & Silvia Ballarè. 2023. Analyzing language variation: where sociolinguistics and linguistic typology meet. In Silvia Ballarè & Guglielmo Inglese (eds.), *Sociolinguistic and typological perspectives on language variation*, 1-28. Berlin: Mouton De Gruyter.
- itTenTen20 corpus, available on <https://www.sketchengine.eu/ittenten-italian-corpus/>
- Keenan, Edward L. & Bernard Comrie. 1977. Non phrase accessibility and universal grammar. *Linguistic inquiry* 8. 63-99.
- Kidd, Evan (ed.). 2011. *The acquisition of relative clauses: processing, typology and function*. Amsterdam/Philadelphia: John Benjamins.
- Labov, William. 1984. Field methods of the project on linguistic change and variation. In John Baugh & Joeal Scherzer (eds.), *Language in use: readings in sociolinguistics*, 28-54. Englewood Cliffs (NJ): Prentice Hall.



- Ledgeway, Adam. 2016. The dialects of southern Italy. In Adam Ledgeway & Martin Maiden (eds.), *The Oxford Guide to the Romance languages*, 246-269. Oxford: Oxford University Press.
- Ledgeway, Adam & Martin Maiden (eds.). 2016. *The Oxford Guide to the Romance languages*. Oxford: Oxford University Press.
- Lepschy, Giulio. 2002. Popular Italian: Fact or Fiction?. In Giulio C. Lepschy (ed.), *Mother Tongues and Other Reflections on the Italian Language*, 49-69. Toronto: University of Toronto Press.
- Levshina, Natalia. 2015. *How to do Linguistics with R*. Amsterdam: John Benjamins.
- Mauri, Caterina, Silvia Ballarè, Eugenio Gorla, Massimo Cerruti & Francesco Suriano. 2019. KIParla corpus: a new resource for spoken Italian. In Raffaella Bernardi, Roberto Navigli & Giovanni Semeraro (eds.). *Proceedings of the 6th Italian conference on Computational Linguistics CLiC-it*, Torino, Accademia University Press.
- Murelli, Adriano. 2011. *Relative constructions in European non-standard varieties*. Berlin/New York: Mouton de Gruyter.
- Renzi, Lorenzo. 2000. Le tendenze dell'italiano contemporaneo. Note sul cambiamento linguistico nel breve periodo. *Studi di lessicografia italiana XVII*. 279-319.
- Renzi, Lorenzo. 2012. *Come cambia la lingua. L'italiano in movimento*. Bologna: Il Mulino.
- Serianni, Luca. 2010 [1989]. *Grammatica italiana: italiano comune e lingua letteraria. Suoni, forme, costrutti*. Torino: UTET.
- Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests and trees of York English: Was/were variation as case study for statistical practice. *Language variation and change* 24(2). 135-178.

#### CONTACT

[silvia.ballare@unibo.it](mailto:silvia.ballare@unibo.it)