

Register impacts syntax: Scaling the accessibility hierarchy of relatives

Silvia Ballarè,^a Pierre Larrivé^b

^a University of Bologna, Italy <silvia.ballare@unibo.it>

^b Université de Caen, Normandie Université, France <pierre.larrivee@unicaen.fr>

This article is concerned with the impact of register on syntax. It considers the well-established syntactic regularity of the accessibility hierarchy of relative pronouns (Keenan & Comrie 1977). Relative pronouns with more accessible functions are reported to have a greater quantitative representation in usage (Biber *et al.* 1999: 609-612). Recent corpus research suggests that this tendency is maintained across registers (Hirschberg, Fery & Roth 2014 for German, Wiechman 2014 for English), but that the dominant proportion of more accessible functions is scaled down in the vernacular (Larrivé & Skrovec 2019 for French). Such an observation is counter-intuitive since the vernacular would be expected to give rise to fewer less-accessible functions. An investigation is therefore conducted to test the scaling effect of register that compares the quantitative representation of each function of relative pronouns in exchanges from vernacular and normative registers of contemporary spoken Italian. The comparison confirms that register scales the hierarchy: the proportion of accessible functions is significantly reduced in the vernacular as compared to normative exchanges. The scaling effect is thus proposed to be a vernacular universal.*

KEYWORDS: accessibility hierarchy, register, relatives, vernacular, Italian.

1. Introduction

Linguists generally agree that observable productions are the results of rules internalized by speakers. One major issue is the extent to which these rules should be categorical or variable. The question applies to diachronic, regional and social variation, as well as to structural variation. For instance, relative pronouns are known to be constrained by an accessibility hierarchy. More accessible functions are more frequent in language productions than less accessible functions (e.g. Biber *et al.* 1999: 609-612). The quantitative gap between the more and the less accessible functions is sensitive to register across languages according to recent work. Results from Wiechman (2014) for English, Hirschberg

* This paper is the result of systematic collaboration between the two authors. Silvia Ballarè wrote sections 3 and 4, while Pierre Larrivé wrote section 2, and sections 1 and 5 were co-written.

et al. (2014) for German and Larrivé & Skrovec (2019) for French, discussed in section 2 below, converge in indicating that, while the hierarchy is unaltered, the quantitative gap between subject, object and other relatives is lesser in the vernacular register of unprepared, unmonitored exchanges, than it is in normative registers.¹ If they were confirmed, these results would have an important consequence, suggesting that grammatical constraints are impacted upon by third factors like register. The assumed dichotomy between structure-internal and structure-external dimensions of language might thus be less absolute than assumed.

The purpose of this work is to test suggestions that register scales the quantitative representation of relative pronoun functions following the accessibility hierarchy.² To do so, a systematic comparative investigation is conducted on the quantitative representation of the functions of relative pronouns in vernacular and normative contemporary Italian exchanges. First, we provide background on the accessibility hierarchy of relatives, and report on recent research suggesting that the proportion of relative pronoun functions is scaled by register: the less normative the register is, the less quantitatively dominant the accessible functions are. Then, we present the protocol by which the correlation is verified. Confirmation obtains that register scales down the proportion of more accessible functions. The scaling effect is proposed to be a vernacular universal.

2. Accessibility Hierarchy

The accessibility hierarchy is a typological hypothesis proposed by Edward Keenan and Bernard Comrie in 1977. It is developed from the study of the shape of relative subordinates across languages. The starting point is the observation that, at least in some languages, these subordinates are headed by a marker that varies in form according to the function that it occupies in the subordinate – which is not typically the case with conditionals or concessives, for example. Such a situation can be illustrated by French.

(1) French

- a. *Le café qui cherche des serveurs*
ART.DEF café REL.NOM search.PRS.3SG PART.PL waiter.PL
'The café that is looking for waiters.'
- b. *Le café que cherchent les clients*
ART.DEF café REL.ACC search.PRS.3PL ART.DEF.PL client.PL
'The café that the clients are looking for.'
- c. *Le café dont les clients parlent*
ART.DEF café REL.IO ART.DEF.PL client.PL talk.PRS.3PL
'The café that the clients are talking about.'

Register impacts syntax: Scaling the accessibility hierarchy of relatives

- d. *Le café où les clients cherchent à se réchauffer*
 ART.DEF café REL.OBL ART.DEF.PL client.PL search.PRS.3PL to REFL warm_up.INF
 'The café where the clients are trying to warm themselves up.'
- e. *Le café dont les clients cherchent à se réchauffer*
 ART.DEF café REL.GEN ART.DEF.PL client.PL search.PRS.3PL to REFL warm_up.INF
 'The café, the clients of which are trying to warm themselves up.'

Each relative pronoun marks the function that the antecedent would perform in the subordinate, subject *qui* (1a), object *que* (1b), indirect object *dont* (1c), locative *où* (1d) and genitive *dont* (1e). That is the order in which Keenan and Comrie place the functions of the relatives that define a relationship of implication between them. This relationship is proposed under the term of an accessibility hierarchy,³ where each function is more accessible than those that follow it, and less accessible than those that precede it.

(2) Accessibility Hierarchy:

SU > DO > IO > OBL > GEN > OCOMP

Thus, the relative with a subject function is the most accessible, more than that of object, and so on. This hierarchy is verified in its effect across languages. Languages that have a relative construction for a less accessible function have one for the more accessible functions, but not the other way around: thus, the least accessible function in the hierarchy, that of comparative relative, represented in Urhobo (Keenan & Comrie 1975: 75, their example (29)),

(3) Urhobo

oshale na l- i Mary rho n- o
 man the REL Mary big than him
 'The man that Mary is bigger than'

as well as in Italian,

(4) Italian

L' uomo del quale / di cui Maria è più alta
 ART.DEF man of.ART.DEF REL of REL Mary be.PRS.3SG more high
 'The man that Mary is bigger than'

is not attested in English or French:

(5) English

?* *The man that Mary is bigger than*

(6) French

**L' homme que Marie est plus grand que*
 ART.DEF man REL Mary be.PRS.3SG more high CONJ
 'The man that Mary is bigger than'

The study by Keenan and Comrie is certainly one of the most successful proposals in modern linguistics. Credited with 3466 citations by Google Scholar in March 2020, its influence goes far beyond typological generalizations. It has implications for language processing, language acquisition and language variation.

If there is such a thing as a hierarchy of functions for relative pronouns, the expectation is that it should have an impact on language processing.⁴ It should be easier and faster to process and produce a subject relative pronoun than an object relative pronoun, as suggested by the studies reviewed in Scontras *et al.* (2015). More accessible options should give way to fewer performance errors. This is found to be the case in vernacular French exchanges (Larrivé & Skrovec 2016). In a second language, more accessible functions should be learned earlier, and give way to fewer performance errors (see Larrivé 2018 and references therein). These expectations have not been confirmed by all studies. The publication by Guasti, Stavrakaki & Arosio (2012) shows that the morphological information provided by Greek relatives influences their acquisition rate. Far Eastern languages without relative movement (Mandarin Chinese, Japanese, Korean) also contradict the predictions of the hierarchy (see Lin 2015 and some of the contributions to Kidd 2011 for example). Despite these restrictions, it can be assumed that the hierarchy of accessibility remains an operating principle: as Behney and Gass note, there is “a clear preference in production and learning order that reflects the A(ccessibility) H(ierarchy)” (2013: 56).

The impact of the accessibility hierarchy on language variation has not been explored as much as could be expected. Some research comes from the L2 acquisition field: Pavesi (1986) is concerned with production of relatives in naturalistic and guided learning settings; and Roberts (2017) proposes a quantitative study of relatives in academic writing by Swedish learners of English, which at no stage depart from the hierarchy. Some observations have been put forward regarding social (Levey 2006) and regional varieties (van Rooy *et al.* 2010, Suarez-Gomez 2014). Yet, relatively little is known about relatives and register (but see Hendery 2012). A relation is suggested between genre of spoken discourse and distribution of individual relatives in the ICE-GB corpus by Eunyong & Kyungyul (2015). The relations between language change and the accessibility hierarchy remains to be explored. The expectation would be that changing relatives would tend to acquire more accessible functions rather than less accessible ones. This is what Ballarè & Inglese (2021) show in the acquisition by Italian *dove* ‘where’ of subject functions, although the evolution jumps from locative to subject and does not go through the intermediate direct and indirect object functions.

Systematic analysis of the relation between register and the accessibility hierarchy of relative pronoun functions is provided by two recent studies. The work by Hirschberg, Fery & Roth (2014) compares among other factors the proportion of relative marker functions in two types of German spoken data: one is the read version of a novel (*Vorleser*), the other is spoken exchanges (*Linderstraße*). The results are provided by the following table. They are statistically significant (at $X^2(df=2, N=1164) = 13.648, p < 0.01$).

RELATIVE PRONOUN	LINDERSTRASSE	VORLESER
SUBJ	20 (25%)	205 (41%)
DO	16 (20%)	127 (25%)
Other	44 (55%)	170 (34%)
Total	80 (100%)	502 (100%)

Table 1. Relative pronouns according to grammatical function.

Larrivée & Skrovec (2019) consider whether the accessibility hierarchy found in normative French holds in the vernacular. The distribution of relative pronouns is investigated in a subset of the ESLO corpus, focusing on five types of oral interactions representing an increasing degree of the vernacular. Degrees of the vernacular are measured following Koch & Oesterreicher (2001), by which the greatest degree of closeness is found in the 24-hour sub-corpus, which are the self-recorded productions of a student over the course of a day, and diminishes with family meals, semi-directed interviews, general assemblies and lectures. It is noted that the quantitative over-representation of more accessible functions is scaled down as the register is more vernacular. The results are summarized in their table 2.

	24H		FAMILY MEALS		SEMI-DIRECTED INTERVIEWS		GENERAL ASSEMBLIES		LECTURES		TOTAL
<i>Qui</i>	123	65,1%	189	70,8%	242	70,1%	456	77%	495	78,9%	1505
<i>Que</i>	53	28%	53	19,9%	61	17,7%	85	14,4%	65	10,4%	317
<i>Où</i>	11	5,8%	22	8,2%	35	1%	36	6,1%	36	5,7%	140
<i>Dont</i>	1	0,5%	3	1,1%	3	0,9%	10	1,7%	21	3,3%	38
<i>Lequel</i> and compounds	1	0,5%	0	0%	4	1,2%	5	0,8%	10	1,6%	20
Total	189	100%	267	100%	345	100%	592	100%	627	100%	2020

Table 2. Distribution of relative pronouns in the ESLO corpus.

As the exchange is less vernacular, the proportion of subject relative *qui* increases progressively from 65% to 79% of all relatives; while the share of direct object *que* decreases from 28% to 10%. Rarer relatives *dont* and *lequel* go from 0,5% to 3,3% and 1,6% respectively. Locative *où* displays no particular pattern, probably due to the lack of a sufficient number of occurrences. Furthermore, the distribution of the relative pronouns between the 24h and the Lectures sub-corpora is statistically significant (at X^2 (df=4, N = 816) = 40.573, $p < 0.01$), although it may not be between ‘closer’ sub-corpora.

Apart from the small number of rarer pronouns, the method could be improved, since the focus is on the forms, and while there is a relatively close association between French relative pronouns and their functions, no distinction is made as to the function of *où* and *dont*, which can be either indirect objects or obliques. Nonetheless, for items with more than 50 attestations, there remains a regular variation between the respective weight of the hierarchy function markers and the proximity of the exchange, such that the more vernacular the exchange, the lesser the dominance of the accessible subject function.

This generalization is confirmed by two studies with more detailed quantitative methodology. Roland, Dick & Elman (2007) contrast types of subject and object relatives across a variety of corpora to provide a benchmark for psycholinguistic studies. Their table 3.

The table shows a disparity as to the relation between subject and object relatives in the written (BNC, Brown, Wall Street) and spoken registers (BNC Spoken, Switchboard). In the former, the ratio of subject to object in the two first cells are 4,8, 7,6 and 10,1 respectively, whereas in the latter it is 2,5 and 1,7. However, the other functions are not accounted for.

An even more exhaustive analysis of the impact of register is provided by Wiechman in his extensive monograph on the behavior of relatives in English (2014). The functions of relatives in each of the written and spoken register is quantified in the following table, adapted from his Figure 32 (p. 104).

Subject relatives consist of 212 occurrences in the spoken and 393 in the written out of 495 and 500 analyzed cases. The disparity between subject and all the other functions in the coded relatives is manifestly greater in the written, where it represents nearly four fifths of all occurrences, as compared to the spoken, where it is almost half that proportion at a little more than two fifths of all occurrences. It should be noted that the distribution of the syntactic functions between the two corpora is statistically significant (at X^2 (df= 3, N=995) = 164.54, $p < 0.01$).

Register impacts syntax: Scaling the accessibility hierarchy of relatives

TYPE OF RELATIVE CLAUSE	BRITISH NATIONAL CORPUS	BRITISH NATIONAL CORPUS SPOKEN	BROWN	SWITCHBOARD	WALL STREET JOURNAL TREEBANK 2
SUBJ rel.	14.182	9851	15.024	9548	18.229
OBJ rel.	2943	3863	1976	5616	1802
OBJ rel. (reduced)	5455	14.423	4746	5314	3385
Passive rel.	3118	1729	2867	302	1224
Passive rel. (reduced)	10.730	2886	10.733	779	12.788
SUBJ infinitive rel.	4726	1808	4245	2425	7109
OBJ infinitive rel.	3229	2271	2139	1369	1831
Passive infinitive rel.	542	185	488	88	421
Total	44.924	37.015	42.218	25440	46.788

Table 3. Frequencies of each type of relative clause per 1 million noun phrases.

RELATIVE PRONOUN	SPOKEN	WRITTEN
SUBJ	212	393
DO	211	44
OBL	68	61
GEN	4	2
Total	495	500

Table 4. Relativized roles across modalities.

This conforms to what had been found in the Longman grammar of spoken and written English. Looking at the function of relative markers, it reports that *Conversation/fiction* have 55% subject gaps, while *News and academic writing* have 75% subject gaps (Biber *et al.* 1999: 609-612). This finding, based on a large corpus of everyday spoken and written registers, is entirely in line with what has been independently shown by

Hirschberg, Fery & Roth (2014), Larrivé & Skrovec (2019), Roland *et al.* (2007) and Wiechman (2014)

The scaling of the quantitative representation of the accessible subject function of relative pronouns across registers is a fascinating observation that indicates an impact of register on syntactic behavior. The observation runs counter to expectations: vernacular production should contain more syntactically simple functions, not less. The unexpectedness of the results calls for verification of its validity. In order to provide such a verification, we set a systematic investigation, the protocol of which is described in the next section.

3. Data and methods

The aim of this study is to compare how relative clauses behave in different diaphasic varieties. In particular, it investigates whether as suggested by previous studies, there is any significant quantitative difference in the representation of each function of relative markers on the accessibility hierarchy in vernacular and normative registers. To this end, we elected to work on Italian. Italian has a system of relative pronouns that display polyfunctionality, by which most engage in more than one syntactic function. The language has not been extensively analyzed for register,⁵ and has resources for this analysis to be performed. The behavior of Italian relatives was investigated in the *KIParla* corpus (Mauri *et al.* 2019). The corpus consists in 661.177 tokens (around 70 hours) of recorded speech collected during different types of interactions within the university setting. The recordings took place at the universities of Torino and Bologna and include several student-student and student-professor interactions occurring in 5 types of situation, namely: (i) lessons, (ii) exams, (iii) office hours, (iv) semi-structured interviews and (v) spontaneous conversations. They thus represent register variation along three dimensions: symmetrical *vs* asymmetrical relations between the speakers; presence *vs* absence of fixed constraints in turn-taking; presence *vs* absence of a fixed topic.

For the aims of our study, the type of relation between the speakers was considered the most relevant dimension to distinguish between formal and informal situations. Hence, formal contexts are the ones where the speakers have an asymmetrical relation in the academic setting, corresponding to professor-student interactions; conversely, informal contexts involve symmetrical relations between the speakers, as is the case in interactions between students. This is synthesized in table 5 below.

REGISTER	TYPE OF RELATION	SPEAKERS	CONTEXTS
Vernacular	Symmetrical	Student-student	Interviews, spontaneous conversation
Normative	Asymmetrical	Student-professor	Exams, lessons, office hours

Table 5. Registers of interaction in the *KIParla* corpus.

Therefore, lessons, exams and office hours were considered formal because of the asymmetrical relation between the speakers (i.e. professors/students); while semi-structured interviews and spontaneous conversations were considered informal, because only students were involved in these interactions. Support for this distinction is brought by the behavior of allocutive pronouns (see Helmbrecht 2013, Coveney 2016, among others): in formal contexts, speakers use the *lei* strategy, while in the informal ones they use the *tu* strategy almost categorically.⁶ The two sub-corpora have a comparable size, the first consisting of 339.752 tokens, the second yielding a total amount of 321.425 tokens.

The *KIParla* corpus consists in raw texts, that are not pos-tagged; for this reason, semi-automatic search was applied to the relative pronouns.⁷ The forms involved in Italian relativization strategies are presented in table 6 (see reference grammar by Serianni 2010 [1989] V, §217-240 for a detailed discussion). The categories used are those from Keenan & Comrie (1997).

FUNCTIONS	STRATEGIES	
	GENDER/NUMBER AGREEMENT	INVARIABLE
SUBJ	<i>il quale</i>	<i>che</i>
DO	<i>il quale</i>	<i>che</i>
IO	PREP. + <i>il quale</i>	PREP. + <i>cui</i> <i>cui</i>
OBL	PREP. + <i>il quale</i>	PREP. + <i>cui</i> <i>dove</i>
GEN	PREP. + <i>il quale</i>	PREP. + <i>cui</i> <i>cui</i>
OCOMP	PREP. + <i>il quale</i>	PREP. + <i>cui</i>

Table 6. Relative pronouns in standard Italian.

Subjects and direct objects can be relativized with *il quale* and *che*. According to grammatical descriptions, the former characterizes formal and written productions, while the latter is by far the most common relativization strategy. All the other syntactic functions can be expressed with *il quale* and *cui*⁸ preceded by a preposition. Sporadically, the bare *cui* relativizes a genitive or an indirect object. Finally, *dove* introduces relative clauses with a locative value.⁹

For the purpose of our study, all the occurrences of the relative pronouns that appear less than 500 times in each of the two register sub-corpora were extracted. This concerned *quale*, *dove* in both sub-corpora and *cui* in normative productions. If the number of occurrences was higher, we extracted the first 500 occurrences per sub-corpus. That was the case for *che* in both sub-corpora, and *cui* in vernacular productions – their token frequencies are marked with an asterisk in the table below. Then, we manually selected just the cases in which the considered element behaves as a relativizer. That is because *cui* and, especially, *che* are polyfunctional markers that can also be employed as complementizers (on the so called ‘*che* polivalente’, i.e. ‘multifunctional *che*’, see Fiorentino 1999, 2007a,b, 2010). The number of selected occurrences is summarized in table 7.

RELATIVE PRONOUN	VERNACULAR	NORMATIVE
(PREP.) + <i>il quale</i>	12	109
<i>Che</i>	204*	288*
(Prep.) + <i>cui</i>	354*	410
<i>Dove</i>	241	119
Total		1737

Table 7. Number of relative pronouns extracted from the *KIParla* register sub-corpora.

Each extracted occurrence was annotated for register (i.e. vernacular or normative) and syntactic function (i.e. SUBJ, DO, IO, OBL, GEN, OCOMP). Each function is illustrated by uses of Italian relatives, *che*, *quale*, *cui*, *dove*. Examples of the items employed to relativize subject (7), direct (8) and indirect object (9), oblique (10) and genitive (11) are provided below.

- (7) [...] *anche nelle sue lettere a Gandusio il quale non*
 also in.ART.DEF.PL his.PL letter.PL to G. ART.DEF REL NEG
capisce niente di tutto ciò
 understand.PRS.3SG nothing of all this
 ‘[...] also in his letters to Gandusio, who does not understand anything about all this.’

Register impacts syntax: Scaling the accessibility hierarchy of relatives

- (8) *Avete visto gli incensi che ho comprato*
 have.PRS.2PL see.PTCP ART.DEF.PL incense.PL REL have.PRS.1SG buy.PTCP
 ‘You saw the incense sticks that I bought.’
- (9) *L’ ultima parte del corso [...] alla quale dedicherò le ultime tre lezioni*
 ART.DEF last part of.ART.DEF course to.ART.DEF REL devote.FUT.1SG ART.DEF
 last three lesson.PL
 ‘The last part of the course [...] to which I will devote the last three lessons.’
- (10) *Li c’ è un bar bellissimo dove al tramonto si vede tutto*
 there there be.PRS.3SG ART.INDF bar beautiful REL
 at.ART.DEF sunset IMPR see.PRS.3SG everything
 ‘Over there, there is a very beautiful bar from where at sunset you can see everything.’
- (11) *[...] movimento del Sessantotto di cui si celebrerà il cinquantesimo anniversario tra due anni*
 movement of.ART.DEF sixty_eight of REL IMPR celebrate.FUT.3SG
 ART.DEF fiftieth anniversary in two years
 ‘[...] Sixty-eight movement of which we will celebrate the fiftieth anniversary in two years.’

The resulting proportion of relatives for each function in each register was then compared to answer the question of whether register scales the proportion of accessible relatives, and specifically the subject function. The results of the comparisons are discussed in the next section.

4. Results

Let us consider the quantitative representation of functions of relative pronouns. Remember that two of the pronouns are represented by more than 500 occurrences, *che* in normative and vernacular sub-corpora, and *cui* in the vernacular sub-corpora. This means that the proportion of their functions in the analyzed sample of 500 occurrences is the primary basis for the estimation of the total number of functions in each sub-corpora. To illustrate, we project 1823 subject uses for *che*; this is obtained by finding 121 subject *che* in the sample of 500 occurrences, for a proportion of 24,2%, which is then applied to the total number of 7535 occurrences. The category ‘other’ in the table refers to other functions of the items; for instance, *che*, being highly polyfunctional, can introduce, among others, arguments and causal clauses (for an overview on the uses of *che* in different varieties of Italian, see Fiorentino 2010).

The results of the projection for these two items are provided in the following table.

- VERNACULAR SUB-CORPUS

Che (total number of occurrences = 7535)

	SUBJ	DO	IO	OBL	GEN	OTHER	TOTAL
Analyzed sample	121	71	0	12	0	296	500
Projection	1823	1070	0	181	0	4461	7535

- NORMATIVE SUB-CORPUS

Che (total number of occurrences = 8580)

	SUBJ	DO	IO	OBL	GEN	OTHER	TOTAL
Analyzed sample	200	87	0	1	0	212	500
Projection	3432	1493	0	17	0	3638	8580

Cui (total number of occurrences = 714)

	SUBJ	DO	IO	OBL	GEN	OTHER	TOTAL
Analyzed sample	0	0	20	369	20	91	500
Projection	0	0	29	527	29	129	714

Table 8. Projected numbers of the functions of relative pronouns with more than 500 occurrences in the *KIParla* register sub-corpora.

This allows the number of occurrences for each function of each item in the two sub-corpora to be provided. They are listed in table 9, where projected numbers are followed by an asterisk, and where proportions are approximated to the second digit.

Register impacts syntax: Scaling the accessibility hierarchy of relatives

	SUBJ	DO	IO	OBL	GEN	TOTAL
VERNACULAR						
<i>Che</i>	1823*	1070*		181*		3074
(prep +) <i>Quale</i>	1			9		10
(prep +) <i>Cui</i>			13	327	15	355
<i>Dove</i>				254		254
Total	1824* (49,40%)	1070* (28,97%)	13 (0,35%)	771* (20,87%)	15 (0,41%)	3693
NORMATIVE						
<i>Che</i>	3432*	1493*		17*		4942
(prep +) <i>Quale</i>	12		2	82	13	109
(prep +) <i>Cui</i>			29*	527*	29*	585
<i>Dove</i>				106		106
Total	3444* (59,99%)	1493* (25,99%)	31* (0,54%)	732* (12,75%)	42* (0,73%)	5742

Table 9. Numbers of the functions of relative pronouns in the *KIParla* register sub-corpora.

Looking at the totals in each sub-corpus, we note that the subject is the most frequently expressed of all functions in the vernacular and normative sub-corpora, with respectively 49% and 60%. Next, we find the direct object function, with 29% and 26% respectively. In both registers, subject and object relative pronouns are by far the most frequent (vernacular sub-corpus: 78%; normative sub-corpus: 86%) while the other three (namely, IO, OBL and GEN) are much less so. Unexpectedly, the next better represented function is not the indirect object, which has very few occurrences, with less than 1%, but the oblique, at 21% and 13% respectively.¹⁰ The genitive is represented at less than 1%. Apart from the gross under-representation of indirect objects, these results follow the predictions of the accessibility hierarchy. In order to verify the correlation between the distribution of relativized syntactic functions and diaphasic variation, we conducted a chi-squared test. The result proved the distribution to be highly statistically significant (with $X^2(df=4, N=9435) = 151.14, p < 0.01$). They also are in line to the results found in Biber *et al.* (1999), Roland, Dick & Elman (2007), Wiechman (2014) and Larrivé & Skrovec (2019).

These data allow us to answer the question of whether there is a difference in the proportion of more accessible functions in each register. First, we note that in the vernacular sub-corpus, there are fewer relative pronouns than in the normative one. This is established not only considering the absolute numbers – 3694 in the vernacular and 5741 in the normative – but also calculating the ratio between the observed frequency of relative clauses and the total number of tokens in each sub-corpus (vernacular: 0,01; normative: 0,02). Such a relation is also the case in the previous studies where a closed corpus of comparable word numbers is examined (thus excluding Wiechman, who looks at 500 occurrences from each register).

	FRENCH – LARRIVÉE & SKROVEC (2019)		GERMAN – HIRSCHBERG, FERY & ROTH (2014)		ITALIAN – KIPARLA CORPUS	
	Most Vernacular	Most Normative	Vernacular	Normative	Vernacular	Normative
Total number of relative pronouns	189	627	80	502	3694	5741

Table 10. Total number of relative pronouns in the vernacular and normative registers.

Register also appears to impact the relative proportion of more accessible functions. Previous studies suggest that the proportion of subject relatives is less dominant in the vernacular than in the normative register.

	ENGLISH – WIECHMAN (2014)		FRENCH – LARRIVÉE & SKROVEC (2019)		GERMAN – HIRSCHBERG, FERY & ROTH (2014)	
	Spoken	Written	Most Vernacular	Most Normative	Vernacular	Normative
SUBJ	42	79	65	79	25	41
OBJ	42	9	28	10	20	25
OTHER	16	12	7	11	55	34

Table 11. Proportion of functions of relative pronouns in the vernacular and normative registers.

Going back to the *KIParla* results presented in Table 9, the proportion of the syntactic functions indeed differs between the two registers:

SUBJ goes from 49,40% in the vernacular productions to 59.99% of the normative ones, OBJ goes from 28,97% to 25,99%, IO and GEN show comparable values (even though their proportion is always higher in the normative productions: 0,35% vs 0,54% and 0,41% vs 0,73%, respectively) while OBL goes from 20,87% to 12,75%. In other words, even though the accessibility hierarchy is followed as expected, what changes is the weight of each function in the two registers: strikingly, the left part of the hierarchy, represented by the more accessible functions, is much 'heavier' in normative productions. Spectacularly, obliques, which are considered to be more difficult to process (Diessel & Tomasello 2005 among others), are more frequent in the vernacular sub-corpus.

<i>KIPARLA</i>		
	Vernacular	Normative
SUBJ	50	60
OBJ	29	26
OTHER	21	14

Table 12. Proportion of functions of relative pronouns in the *KIParla* register sub-corpora.

The results from the *KIParla* register sub-corpora validate the suggestion found in current research: register scales the proportion of accessible functions of relative pronouns. The significance of these results is discussed in the final section.

5. Conclusive discussion

The point of departure of this article is the observed variation between register and frequency of relative pronoun functions. The frequency of relative pronouns has been found to depend on their function along the accessibility hierarchy proposed by Keenan & Comrie (1977). Thus, the subject relative pronouns, which are the most accessible on the hierarchy, are the most frequent. This frequency seems to co-vary with the degree of formality of the exchange: the more normative the language register, the greater the proportion of subjects. This is suggested by recent investigations of English, German and French.

Our objective is to verify these observations with a dedicated methodology. We used the *KIParla* corpus of contemporary Italian realized in a defined university environment that is balanced for register, and

extracted up to 500 occurrences of each marker in each sub-corpus of vernacular and normative productions. The results confirm the observations from previous studies. In vernacular exchanges, the proportion of subject relatives is lower than in the formal exchanges, and so is the combined proportion of subject and object relatives. So, while the quantitative representation respects the accessibility hierarchy in both registers, as shown by basic statistical tests applied to central figures, it is scaled by register both with respect to the proportion of relatives, and the proportion of more accessible function. Given their attestation in existing research, these two observations appear to qualify as what Chambers calls a ‘vernacular universal’ (2000, 2001, 2004).

However, the question arises as to why it should be so. Why should there be fewer relatives in the vernacular? Relatives in the vernacular corpus represents 39% of all relatives in the *KIParla* corpus, a proportion that is 25% in the French corpus study and 14% in the German corpus. This may be attributed to the pressures of online production, that leads speakers to avoid more complex configurations, and hence use fewer relative clauses. As to the lesser proportion of more accessible functions, documented in tables 11 and 12, we offer the following speculation. It might be that some minimum proportion of less accessible functions have to be produced to satisfy subcategorization requirements; verbs calling for an indirect object (like ‘go’ or ‘speak’) may be proportionally more frequent in a relative in the vernacular than in the normative register; it would follow that more accessible functions would be relatively less frequent. The speculation that we formulate relates two cross-linguistically stable behavior of relatives in the vernacular register, a scaling down of more accessible functions and a lower proportion of relatives overall. A way to verify this would be to examine the representation of personal pronouns, and the relative frequency of verbs with indirect objects, and it may be that the use of sophisticated statistical models would help in revealing correlative variation in future research.

Notes

¹ The notion of register is notoriously difficult to define. We use the terms normative and vernacular reflecting the two poles of communicative distance and proximity respectively, referring to the work by Koch & Oesterreicher (e.g. 2001).

² This objective entails that a number of issues regarding relative clauses and relative pronouns fall outside the remit of the work. On syntactic treatments of relative clauses, see e.g. Arnold & Godard (*forthcoming*), Cinque (2020), Bianchi (2011a, b); on functional approaches to relative clause behavior, see Fox & Thompson (1990), Michaelis & Duffield (2001), and Larrivé & Skrovec (2019).

³ Note that ‘accessibility’ is about the “relative accessibility to relativization of NP

positions in simplex main clauses” (Keenan & Comrie 1977: 66).

⁴ For recent studies on processing of relative clauses, see among others Scontras *et al.* (2015), Gibson *et al.* (2013), and contributions in Kidd (2011).

⁵ For some figures of relatives per function in vernacular Italian, see Fiorentino (2007b: 279).

⁶ There is one case in which a professor uses the *tu* strategy while addressing to a student during an exam. Note that this only underlines the asymmetric relation between the two speakers, as the student would not use *tu* to address the professor.

⁷ Although this is largely orthogonal since we are investigating the function of relatives as situated on the Accessibility Hierarchy and not the behavior of the relative clause, we note that, in the corpus, appositive relative clauses are exceedingly rare in the vernacular, as they are in e.g. the French data (Larrivée & Skrovec 2019); we have noted no example in the Italian corpus that requires an appositive analysis.

⁸ Following Serianni (2010 [1989], V, §221), we classify *cui* as invariable, since it cannot be inflected. However, unlike *che* and *dove*, *cui* can be preceded a preposition (expressing its syntactic function) and/or a definite article (that shows gender/number agreement with the head of the relative clause). In other words, considering the relativization strategies identified by Comrie & Kuteva (2013), while *che* and *dove* are to be considered as examples of the ‘gap strategy’ (in that the syntactic function is not expressed), *cui* (as *il quale*), when preceded by a preposition, can be considered as a relative pronoun *stricto sensu* since the syntactic role of the relativized item is overtly expressed.

⁹ Italian displays a complex paradigm of relative pronouns and relative clauses that have been discussed in great detail in the literature from different perspectives (see Cinque 1978, 1988, Bernini 1989, Fiorentino 1999, 2007a,b, Cristofaro & Giacalone Ramat 2002, 2007, Cerruti 2016, 2017, Ballarè, Cerruti & Gorla 2019, among others).

¹⁰ Note that all the cases in which *dove* relativizes a prepositional phrase are coded as OBL. Furthermore, *che* in real corpus data also introduces OBL: in the vast majority of cases, *che* relativizes temporal obliques and this use can be considered standard (see Berruto 2017: 43).

Bibliographical References

- Arnold, Doug & Godard, Danièle *forthcoming*. Relative clauses in HPSG. In Müller, Stefan; Abeillé, Anne; Borsley, Robert D. & Koenig, Jean-Pierre (eds.), *Head-Driven Phrase Structure Grammar: The handbook*. Berlin: Language Science Press.
- Behney, Jennifer & Gass, Susan 2013. Accessibility Hierarchy. A study using syntactic priming. In Schwieter, John W. (ed.), *Innovative research and practices in Second Language Acquisition and bilingualism*. Amsterdam: Benjamins. 43-61.
- Ballarè, Silvia; Cerruti, Massimo & Gorla, Eugenio 2019. Variazione diastratica nel parlato di giovani: il caso delle costruzioni relative. In Moretti, Bruno; Kunz, Aline; Natale, Silvia & Krakenberger, Etna (eds.), *Le tendenze dell’italiano contemporaneo rivisitate. Atti del 52^o Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI)*, Berna, 6-8 settembre 2018. Milano: Officinaventuno. 75-94.
- Ballarè, Silvia & Inglese, Guglielmo 2021. The development of locative relative markers: from typology to sociolinguistics (and back). *Studies in Language – Online-First Article* <www.jbe-platform.com/content/journals/10.1075/

- sl.20013.bal > .
- Bernini, Giuliano 1989. Tipologia delle frasi relative italiane e romanze. In Foresti, Fabio; Rizzi, Elena & Benedini, Paola (eds.), *L'italiano tra le lingue romanze, Atti del 20° Congresso Internazionale di Studi della Società di Linguistica Italiana (SLI)*, Bologna, 25-27 settembre 1986. Roma: Bulzoni. 85-98.
- Berruto, Gaetano 2017. What is changing in Italian today? Phenomena of restandardization in syntax and morphology: an overview. In Cerruti, Massimo; Crocco, Claudia & Marzo, Stefania (eds.), *Towards a new standard. Theoretical and Empirical Studies on the Restandardization of Italian*. Berlin / New York: Mouton De Gruyter. 31-60.
- Bianchi, Valentina 2002a. Headed relative clauses in Generative Syntax. Part I. *Glott International* 6,7. 197-204.
- Bianchi, Valentina 2002b. Headed relative clauses in Generative Syntax. Part II. *Glott International* 6,8. 235-247.
- Biber, Douglas; Johansson, Stig; Leech, Geoffrey; Conrad, Susan & Finegan, Edward 1999. *The Longman grammar of spoken and written English*. London: Longman.
- Cerruti, Massimo 2016. Costruzioni relative in italiano popolare. In Guerini, Federica (ed.), *Italiano e dialetto bresciano in racconti di partigiani*. Roma: Aracne. 79-118.
- Cerruti, Massimo 2017. Changes from below, changes from above: relative constructions in contemporary Italian. In Cerruti, Massimo; Crocco, Claudia & Marzo, Stefania (eds.), *Towards a new standard. Theoretical and Empirical Studies on the Restandardization of Italian*. Berlin / New York: Mouton De Gruyter. 61-88.
- Chambers, Jack 2000. Universal sources of the vernacular. In Ammon, Ulrich; Mattheier, Klaus J. & Nelde, Peter H. (eds.), *The future of European Sociolinguistics* (Special issue of *Sociolinguistica: International Yearbook of European Sociolinguistics* 14). Tübingen: Niemeyer. 11-15.
- Chambers, Jack 2001. Vernacular Universals. In Fontana, Joseph M.; McNally, Louise; Turrel, M. Teresa & Vallduví, Enric (eds.), *ICLaVE1: Proceedings of the First International Conference on Language Variation in Europe*. Barcelona: Universitat Pompeu Fabra. 52-60.
- Chambers, Jack 2004. Dynamic typology and vernacular universals. In Kortmann, Bernd (ed.), *Dialectology meets Typology. Dialectic Grammar from a Cross-Linguistic Perspective*. Berlin / New York: Mouton De Gruyter. 127-146.
- Cinque, Guglielmo 1978. La sintassi dei pronomi relativi cui e quale nell'italiano moderno. *Rivista di grammatica generativa* 3. 31-126.
- Cinque, Guglielmo 1988. La frase relativa. In Renzi, Lorenzo; Salvi, Giampaolo & Cardinaletti, Anna (eds.), *Grande grammatica italiana di consultazione*, I. Bologna: Il Mulino. 443-503.
- Cinque, Guglielmo 2020. *The syntax of relative clauses*. Cambridge: Cambridge University Press.
- Comrie, Bernard 1989. *Language Universals and Linguistic Typology. 2nd edition*. Chicago: The University of Chicago Press.
- Comrie, Bernard & Kuteva, Tania 2013. Relativization strategies. In Dryer, Matthew & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

- Coveney, Aidan 2016. La quête du vernaculaire dans l'étude de la variation grammaticale. *Congrès Mondial de Linguistique Française – CMLF 2016*. Paris: Institut de Linguistique Française.
- Cristofaro, Sonia & Giacalone Ramat, Anna 2002. Relativization patterns in Mediterranean languages, with particular reference to the relativization of circumstantials. In Ramat, Paolo & Stolz, Thomas (eds.), *Mediterranean languages*. Papers from the Med-Typ Workshop (Tirrenia, 1-3 June, 2000). Bochum: Brockmeyer. 99-112.
- Cristofaro, Sonia & Giacalone Ramat, Anna 2007. Relativization strategies in the languages of Europe. In Ramat, Paolo & Roma, Elisa (eds.), *Europe and the Mediterranean as linguistic areas: Convergences from a historical and typological perspective*. Amsterdam / Philadelphia: Benjamins. 63-93.
- Diessel, Holger & Tomasello, Michael 2005. A new look at the acquisition of relative clauses. *Language* 81. 1-25.
- Eunyoung, Jang & Kyungyu, Kim 2015. A corpus-based analysis of the variation of relative pronouns depending on genres. *Studies in Linguistics* 34. 261-280.
- Fiorentino, Giuliana 1999. *Relativa debole. Sintassi, uso, storia in italiano*. Milano: Franco Angeli.
- Fiorentino, Giuliana 2007a. Le relative 'pragmatiche' in italiano. In Venier, Federica (ed.), *Relative e pseudo-relative tra grammatica e testo*. Alessandria: Dell'Orso. 53-71.
- Fiorentino, Giuliana 2007b. European relative clauses and the uniqueness of the Relative Pronoun Type. *Italian Journal of Linguistics* 19,2. 263-291.
- Fiorentino, Giuliana 2010. *Che polivalente*. In Simone, Raffaele; Berruto, Gaetano & D'Achille, Paolo (eds.), *Enciclopedia dell'italiano*. Roma: Treccani.
- Fox, Barbara & Thompson, Sandra A. 1990. A discourse explanation of the grammar of relative clauses in English conversation. *Language* 66,2. 297-316.
- Gibson, Edward; Tily, Harry & Fedorenko, Evelina 2013. The processing complexity of English relative clauses. In Sanz, Montserrat; Laka, Itziar & Tanenhaus, Michael K. (eds.), *Language down the garden path: The cognitive and biological basis for linguistic structures*. Oxford: Oxford University Press. 149-173.
- Gordon, Peter C. & Lowder, Matthew W. 2012. Complex Sentence Processing: A Review of Theoretical Perspectives on the Comprehension of Relative Clauses. *Language and Linguistics Compass* 6,7. 403-415.
- Guasti, Maria Teresa; Stavrakaki, Stavroula & Arosio, Fabrizio 2012. Cross-linguistic differences and similarities in the acquisition of relative clauses: Evidence from Greek and Italian. *Lingua* 122, 6. 700-713.
- Helmbrecht, Johannes 2013. Politeness distinctions in pronouns. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Hendery, Rachel 2012. *Relative clauses in time and space: A case study in the methods of diachronic typology*. Amsterdam / Philadelphia: Benjamins.
- Hirschberg, Tim; Fery, Caroline & Roth, Anne 2014. Relative clauses in colloquial and literary German: A contrastive corpus-based study. *Linguistische Berichte* 240. 405-445.
- Keenan, Edward L. & Comrie, Bernard 1977. Noun Phrase accessibility and Universal Grammar. *Linguistic Inquiry* 8,1. 63-99.

- Kidd, Evan (ed.) 2011. *The acquisition of relative clauses. Processing, typology and function*. Amsterdam: Benjamins.
- Koch, Peter & Oesterreicher, Wulf 2001. Gesprochene Sprache und geschriebene Sprache / Langage parlé et langage écrit. In Holtus, Günter; Metzeltin, Michael & Schmitt, Christian (eds.), *Lexikon der Romanistischen Linguistik, Bd. I/2*. Tübingen: Niemeyer. 584-627.
- Larrivée, Pierre 2018. Complexité et relatives en français vernaculaire langue maternelle et langue seconde. In Bang-Nilsen, Catrine (ed.), *Acquisition du français – Synergies Pays Scandinaves* 13, 65-77.
- Larrivée, Pierre & Skrovec, Marie 2019. Variation en français vernaculaire: le cas des relatives. *L'Information grammaticale* 163. 21-28.
- Larrivée, Pierre & Skrovec, Marie 2016. Les relatives en français vernaculaire. Neveu, Franck; Bergounioux, Gabriel; Côté, Marie-Hélène; Fournier, Jean-Marie; Hriba, Linda & Prévost, Sophie (eds.), *Congrès Mondial de Linguistique Française 2016*. SHS Web of Conferences.
- Levey, Stephen 2006. Visiting London relatives. *English World-Wide* 27,1. 45-70.
- Lin, Yow-Yu 2015. Developmental trajectory of relative clauses among Mandarin learners of English. *Concentric: Studies in Linguistics* 41,1. 63-94.
- Mauri, Caterina; Ballarè, Silvia; Gorla, Eugenio; Cerruti, Massimo & Suriano, Francesco 2019. KIParla corpus: A new resource for spoken Italian. In Bernardi, Raffaella; Navigli, Roberto & Semeraro, Giovanni (eds.), *Proceedings of the 6th Italian conference on Computational Linguistics CLiC-it*.
- Michaelis, Laura A. & Duffield, Cecily Jill 2011. Why subject relatives prevail. Constraints versus constructional licensing. *Language and Cognition* 3,2. 171-208.
- Pavesi, Maria 1986. Markedness, discursal modes, and relative clause formation in a formal and an informal context. *Studies in Second Language Acquisition* 8,1. 38-55.
- Roberts, Tim 2017. *Relativisation in the writing of Swedish L2 Learners of English*. MA dissertation. Uppsala Universiteit.
- Roland, Douglas; Dick, Frederic & Elman, Jeffrey L. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57,3. 348-379.
- van Rooy, Bertus; Terblanche, Lize; Haase, Cristoph & Schmied, Joseph 2010. Register differentiation in East African English: A multidimensional study. *English World-Wide* 31,3. 311-349.
- Scontras, Gregory; Badecker, William; Shank, Lisa; Lim, Eunice & Fedorenko, Evelina 2015. Syntactic complexity effects in sentence production. *Cognitive Science* 39,3. 559-583.
- Serianni, Luca 2010 [1989]. *Grammatica italiana. Italiano comune e lingua letteraria*. Torino, UTET.
- Suarez-Gomez, Cristina 2014. Relative clauses in Southeast Asian Englishes. *Journal of English Linguistics* 42,3, 245-268.
- Wiechmann, Daniel 2014. *Understanding relative clauses. A usage-based view on the processing of complex constructions*. Berlin: De Gruyter Mouton.