

MULTIFUNCTIONAL *CHE*, CLEFT SENTENCES AND PRESENTATIONAL *C'È* IN CORPORA OF SPOKEN ITALIAN

PATRIZIA GIAMPIERI

University of Perugia

pgiampieri@libero.it

Abstract

This paper is aimed at exploring some traits of neo-standard Italian in four different spoken corpora, such as children's conversations, adult conversations, parliamentary debates, and film subtitles. The following features are investigated: multifunctional *che*, cleft sentences, and presentational *c'è*. Building on scholarly examples and sample phrases containing the traits above mentioned, this paper discusses how to search for, retrieve and analyse neo-standard phrasemes in the four corpora. Other features are also presented and investigated on a case-by-case basis. In addition, relative frequencies of phraseme occurrences are noticed and discussed. In this way, a qualitative and a small-scale quantitative comparative approach is developed. The paper findings report that all the traits appear in the corpora with some notable genre-related differences.

Keywords: neo-standard Italian; corpus analysis; spoken language; the grammar of spoken Italian; spoken Italian corpora

1. Introduction

According to Berruto (1987 and 2017, p. 33), neo-standard Italian is colloquial language in formal and informal circumstances, used by both educated and non-educated people. Sabatini (1985, p. 17) defines it as *italiano dell'uso medio parlato e scritto* ("average spoken and written Italian") and as a variety of national language. Ballarè (2020, p. 476) argues that the traits of neo-standard Italian are more similar to spoken than written language.

The acknowledgement of the features of spoken Italian is a recent phenomenon, which dates back to the second half of the 20th century. Before that, people used dialects in everyday conversations. Voghera (1992, p. 60) claims that neo-standard Italian has progressively taken the place of both dialects and formal Italian.

1.1 Traits of neo-standard Italian

The characteristics of neo-standard Italian have been analysed and discussed at length in the literature (Sabatini 1985; Berruto 1987 and 2017; Alfonzetti 2002; D'Achille 2019; Author 2025). Amongst others, they are the following ones: multifunctional *che* (Berruto 2017, p. 43; Poletto and Sanfelici 2017); substandard relative clauses (Alfonzetti 2002; Poletto and Sanfelici 2017); cleft sentences (Sabatini 1985; Berruto 1987 and 2017); presentational *c'è* (Berruto 1985, p. 128 and 2017, pp. 39-41); anacolutha (Sabatini 1985; Alfonzetti 2002; Menin 2014); actualising *ci* (Sabatini 1985, p. 161; Berruto 2017); left and right dislocations (Berruto 1985, pp. 126-127; Gavazzi 2009; D'Achille 2019; Chekalil 2020); duplicated deixis (Berruto 1985), and replacement of the subjunctive mode with the indicative one (Sabatini 1985, pp. 166-167; Alfonzetti, 2002, p. 17).

In multifunctional *che* (“that”) phrases, the *che* connective is used for a variety of syntactical purposes. In particular, it may serve a temporal function (instead of using *quando*, “when”) (example: *l'anno che sono nato*, “the year that I was born”) (Poletto and Sanfelici 2017, p. 807); it is applied to connect sentences (example: *fortuna che mi hai trovato*, “(it was) luck that you found me”) (Sabatini 1985, pp. 164-165); it can be restated (or retained) by a following direct or indirect pronoun (example: *il mio amico che gli parlo*, “my friend that I speak to him”) (Poletto and Sanfelici 2017, p. 807), and it may express causes (example: *aspetta che te lo spiego*, “wait that I’ll explain it to you”) (Sabatini 1985, pp. 164-165).

From the multifunctional *che*, many substandard relative clauses derive. They are defined as such because they are considered as below standard and pertaining to low-level language (Alfonzetti 2002; Berruto 2017; Author 2025). They are subordinate clauses which are always introduced by *che* (“that”) instead of other more formal (and more complex) relative pronouns, such as “of whom/which”, “to whom/which”, etc. (Alfonzetti 2002). Examples are *il professore che conosco bene i suoi lavori* (“the professor that I know his works well”) (Poletto and Sanfelici 2017, p. 807), instead of *il professore di cui conosco bene i lavori* (“the professor whose works I know well”).

Cleft sentences are divided into two parts: the first one containing the finite “be” and the second one introduced by a pseudo-relative clause (Berruto 1987). The relative clause is defined as “pseudo” because syntactically it is not such, as its functions are only to provide new information and prolong the duration of a statement. In such manner, cleft sentences are articulated to foster comprehension (Sabatini 1985, p. 163). An example is *è per il tuo bene che ti parlo* (“it’s for your own good that I’m talking”) (Sabatini 1985, p. 163). Cleft sentences tend to recur in questions starting with *quand’è che*, *dov’è che*, *com’è che* and *chi è che* (ibid.) (“when is it that”, “where is it that”, “how is it that”, and “who is it that”).

A sub-type of cleft sentences are sentences introduced by the presentational *c'è* (Berruto 1985, p. 128 and 2017, pp. 39-41). In this case, *c'è* or *ci sono* (“there is” and “there are”) precede a pseudo-relative clause. An example is in phrase *c'è un gatto che gioca nel giardino* (“there’s a cat that is playing in the garden”), instead of

the standard *un gatto gioca nel giardino* (“a cat is playing in the garden”) (Berruto 1985, p. 128). This type of sentences are aimed at providing new information and giving more emphasis to a particular utterance.

Anacolutha express syntactical inconsistencies (Sabatini 1985; Alfonzetti 2002), where there are no explicit or syntactically coherent connections between two parts of a sentence. An example is the following: *perché una donna sola su una nave è pericoloso* (“because a woman alone on a boat is dangerous”) (Menin 2014, p. 128), instead of *perché è pericoloso per una donna stare sola su una nave* (“because it is dangerous for a woman to be alone on a boat”). As can be seen, *perché una donna sola su una nave* is a syntactically incoherent subject.

The actualising *ci* is a clitic (or redundant) pronoun that does not add any semantic or spatial meaning as it would do in phrases such as *crederci* (“believe in it”) or *c’è* (“there is”) (Sabatini 1985; Berruto 2017, p. 46). An example is *ci vuole pazienza* (“there must be patience”) (Sabatini 1985, p. 161).

Left dislocations are characterised by a clitic pronoun preceding a verb and following phrase elements it refers to (Chekalil 2020, p. 14, 20; D’Achille 2019). An example is *gli amici li hai visti* (Chekalil 2020, p. 22), where the pronoun *li* re-states the noun phrase and phrase subject *gli amici*.

Right dislocations are characterised by a clitic pronoun which anticipates a head noun or a noun phrase functioning as a direct or indirect object (Gavazzi 2009). An example is *li leggo i giornali* (Berruto 2017, p. 40).

Duplicated deixis is due to the egocentric and contextual nature of spoken language. Double deictic elements are, for example, *questo qui* and *quello là* (“this one here” and “that one there”) (Berruto 1985, pp. 126-127).

The subjunctive mode is generally simplified and turned into the indicative one in neo-standard Italian (Sabatini 1985, pp. 166-167). Alfonzetti (2002, p. 17), in fact, refers to a *rarefazione del congiuntivo* (“rarefaction of the subjective mode”). An example is the sentence *pensavo che non c’era bisogno di dirlo* (“I thought there would be no need to say it”), instead of *pensavo che non ci fosse bisogno di dirlo* (“I thought there was/were no need to say it”) (Menin 2014, p. 127).

2. Aim of the paper and research questions

This paper is aimed at exploring some of the above-mentioned traits of neo-standard Italian in corpora of spoken Italian. The traits focused on are as follows: multifunctional *che*, cleft sentences, and presentational *c’è*. The following spoken genres are considered: children’s talk; adult conversations; parliamentary debates, and film transcripts. As can be seen, both naturally occurring and scripted language are addressed. In this way, it is possible to bring to the fore (dis)similarities, if any.

The research question that this paper wishes to address is the following one: “are there any differences in the frequencies of usage and in the types of traits of neo-standard Italian in the spoken corpora investigated in this paper?”.

To answer the question, this paper analyses the relative frequencies of the above-mentioned traits of neo-standard Italian and sources sample phrases from

the corpora of spoken Italian. In such manner, both a qualitative and a small-scale quantitative approach can be developed.

3. Methodology

The corpora consulted for the purposes of this paper are the following ones: Child Language Data Exchange System (CHILDES) (MacWhinney and Snow 1985); Spoken Italian and Who Speaks Italian (KIParla) (Mauri et al. 2019); Italian Parliamentary debates (ParlaMint) (Erjavec et al. 2021), and Open Subtitles 2018 – Italian (Tiedemann 2012). The corpora, with the exception of KIParla¹, are available from the Sketch Engine platform.

CHILDES is composed of transcriptions of spontaneous conversational interactions of young children conversing with their parents or siblings. It contains 459,881 words (572,217 tokens).

KIParla (henceforth abbreviated “KIP”) is a corpus of spoken Italian with more than 150 hours of adult conversations that were recorded, amongst others, in the cities of Bologna and Turin. It comprises 1,385,219 words (1,624,750 tokens).

ParlaMint (abbreviated “PM”) contains transcripts of parliamentary speeches carried out in the Italian Parliament. It comprehends 26,549,927 words (30,722,200 tokens).

The Open Subtitles 2018 corpus (Italian) (“OS” in short) is a collection of original and translated movie subtitles sourced from the opensubtitles.org database. It is composed of 431,415,848 words (573,531,295 tokens).

Some of the traits of neo-standard Italian are investigated in the above-mentioned corpora. In particular, the following ones are focused on: multifunctional *che*, cleft sentences, and presentational *c'è*, along with other ones, as the case may be. For the purpose of the analysis, the following categories of phrasemes are queried: 1) multifunctional *che* phrasemes: *è uno che*, *è una che*, *è quello che*, *è quella che*, *giorno/sera/mese/anno... che* (“he’s one that”, “she’s one that”, “he’s the one that”, “she’s the one that”, “day/evening/month/year... that”); 2) cleft sentences: *è... che*, *chi è che*, *come è che*, *com'è che*, *quando è che*, *quand'è che* (“it’s... that”, “who is that”, “how is that” “how’s that”, “where is that” and “where’s that”), and 3) presentational *c'è* phrasemes: *c'è... che*, *vi è... che*, *v'è... che* (“there is... that” in its informal and formal forms). Other traits of standard and neo-standard Italian are dealt with on a case-by-case basis, as they emerge from the statements or phrases sourced from the corpora.

The above groups of phrasemes have been chosen in light of the examples provided for in the literature, and on the basis of the main characteristics of multifunctional *che* phrases, cleft sentences, and presentational *c'è* phrases discussed above.

In the sections that follow, the phrasemes are searched for in each corpus. To substantiate the analyses and interpretation of the data, relative frequencies and sample phrases (or corpus-driven statements) are reported in two distinct sets of tables.

¹ The KIParla corpus is available at <https://kiparla.it/en/>.

The first set lists relative frequencies, which are calculated per million tokens. For example, if the number of occurrences of the phraseme *è uno/una che* is 2 in CHILDES, its relative frequencies equal 3.5, which is obtained by dividing 2 by 572,217 tokens and by multiplying it by 1,000,000. The second set of tables shows sample phrases (or statements) sourced from each spoken corpus. In this way, small-scale quantitative and qualitative comparative analyses can be carried out.

4. Analysis

This section presents and discusses comparative corpus-assisted analysis of the phrasemes above mentioned.

4.1 Relative frequencies in the spoken corpora

Table 1 reports the relative frequencies of the traits of neo-standard Italian across the spoken corpora. The highest frequencies per phraseme are marked in bold.

Table 1. Relative frequencies of traits of neo-standard Italian across the corpora

Traits and phrasemes		Corpora			
		CHILDES	KIP	PM	OS
Multifunctional <i>che</i>	<i>È uno/una che</i>	3.50	24.62	1.76	3.03
	<i>È quello/quella che</i>	20.97	147.10	61.68	44.82
	<i>Giorno / sera / mese / anno... che</i>	19.22	90.48	137.43*	74.89
Cleft sentences	<i>È... che</i>	145.05	364.98	571.64	354.30
	<i>Chi è / com'è / quand'è... che</i>	135.71	107.71	5.34	14.73
Presentational <i>c'è</i>	<i>C'è... che</i>	249.90	352.05	128.51	108.41
	<i>Vi è... che</i>	-	-	19.21	0.46

**Note: mostly standard usages (see Table 3 for sample phrases)

As can be seen from Table 1, the highest frequencies of multifunctional *che* phrasemes with constructions composed of *è uno/una che* and *è quello/quella che* appear in KIP. Expressions such as *giorno/sera/mese... che* are apparently more recurrent in PM. However, if analysed in context (see Table 3), the majority of such phrasemes are used in a standard way (e.g., *l'ultimo giorno che hanno a disposizione*, “the last day that they have at their disposal”). Therefore, the highest frequencies of neo-standard constructions of the type *giorno/sera/mese... che* are actually found in KIP. In PM, cleft sentences of the type *è... che* prevail. Cleft sentences in questions starting with *chi è che*, *com'è che* and *quand'è che* are the most recurrent in CHILDES. Finally, the more formal *vi è... che* appears the most in PM. As can be seen, OS does not show any particularly high frequencies.

As a whole, it can be seen that expressions such as *chi è che*, *com'è che* and *quand'è che* probably pertain more to children's and spontaneous conversations (see the CHILDES and KIP cells). Conversely, phrasemes such as *vi è... che* are

more recurrent in formal contexts, such as parliamentary debates. The next section clarifies the search syntax used to retrieve the above frequencies.

4.1.1 Corpus search syntax

This section shows the corpus search syntax that was written to obtain both the frequencies and sample phrases of the above phrasemes. In this way, the full replication of the language investigation is ensured.

Table 2. Corpus search syntax per phraseme

No	Phraseme	Search string	With Lemma (optional)
1	<i>È uno / una che</i>	<i>è un* che</i>	-
2	<i>È quello / quella che</i>	<i>è quel* che</i>	-
3	<i>Giorno / sera / mese / anno... che</i>	<i>giorno sera mese anno</i>	<i>che</i> (4 words to the right)
4	<i>È... che</i>	<i>[[[:punct:]] è</i>	<i>che</i> (4 words to the right)
5	<i>Chi è / com'è / quand'è... che</i>	<i>chi è che com'è che come è che quand'è che quando è che</i>	-
6	<i>C'è... che</i>	<i>c'è</i>	<i>che</i> (4 words to the right)
7	<i>Vi è... che</i>	<i>vi è v'è</i>	<i>che</i> (4 words to the right)

In line 1 and line 2, the asterisk in *un** and *quel** allows for the retrieval of any words starting with *un* and *quel*, respectively. In this way, *uno*, *una*, *quello* and *quella* can be obtained. The straight line “|” in line 3 functions as a Boolean OR operator. Therefore, the search for *giorno|sera|mese|anno* together with the word *che* (within a span of 4 words to the right) provides results with *giorno... che*, *sera... che*, *mese... che* or *anno... che*. In line 4, the symbol “[[:punct:]]” represents punctuation marks. In this way, the system searches for any punctuation mark preceding *è... che*. This means that unrelated expressions such as *si è scordato che* (“he has forgotten that”) are eschewed. Finally, in line 5 and line 7, the search for *chi è... che*, *com'è... che*, *quand'è... che* and *vi è... che* considers both contracted and non-contracted forms (e.g., *com'è* and *come è*).

The next sections show sample phrases sourced from the corpora.

4.2 Samples of multifunctional *che* phrases

Table 3 reports samples of multifunctional *che* phrases. There are at least two examples per corpus. Back-translations are written in brackets. Other traits of standard and neo-standard spoken language are marked in bold and discussed below.

Table 3. Samples of multifunctional *che* phrases sourced from the corpora

No.	Corpus / Phraseme	CHILDES	KIP	OS	PM
1	<i>È uno / una che</i>	<p>-Questo che <i>cos'è? È uno che nuota un signore</i> ("what's this? It's one that swims, a man")</p> <p>-E questo è uno che va a letto ("and this is one who goes to bed").</p>	<p>-È una che usa perfettamente il computer ("she's one that uses the computer perfectly")</p> <p>-Il mimmi non è uno che sta con le mani in mano ("mimmi is not one who can sit on his hands")</p> <p>-Tua sorella è una che i plumcake ci stravede ("your sister is one who plum cakes is very fond of them").</p>	<p>-È una che la sa portare la macchina! ("she's one who knows how to drive it, the car")</p> <p>-È uno che nessuno riceve ("he's one who nobody hosts")</p> <p>-Insisto a pensare che è una che beve ("I insist in thinking that she's one who drinks").</p>	<p>-Lei è uno che legge ("you're one who reads")</p> <p>-È uno che dice di essere contro questo prodotto ("he's one who says to be against this product").</p>
2	<i>È quello / quella che</i>	<p>-Non è quello che usi sempre vero? ("isn't it the one that you always use right?")</p> <p>-Cos'è quello che hai in mano Federica? ("what is that you're holding in your hand Federica?").</p>	<p>-Quindi questo è quello che si chiama alla vicentina? ("so this is what is called the vicentina way?")</p> <p>-È quel che tu senti attraverso il naso ("that's what you feel through the nose").</p>	<p>-Questo è quello che sta succedendo ("this is what is happening")</p> <p>-Bon Bon è quella che ti vuole più bene di tutti ("Bon Bon is the one who loves you the most").</p>	<p>-La questione allora è quella che ha posto la senatrice ("the question is then the one posed by the senator")</p> <p>-Perché è quello che chiedono i cittadini ("because this is what citizens ask for").</p>

3	<p><i>Giorno / sera / mese / anno... che</i></p>	<p><i>-Un giorno che lui dormiva</i> ("a day that he was sleeping")</p> <p><i>-Io non c'ero quella sera là che c'era il film</i> ("I wasn't there that evening there that the film was on").</p>	<p><i>-Il primo anno che tu sei venuta a trovarmi</i> ("the first year that you came to visit me")</p> <p><i>-Il giorno che c'era la finale della coppa intercontinentale</i> ("the day that there was the final round of the inter-continental cup")</p> <p><i>-Un bel giorno che la fontana non usciva più acqua</i> ("one fine day that the fountain did not go out any more water").</p>	<p><i>-Ne ho avuto uno il giorno che sei apparso</i> ("I had one the day that you appeared")</p> <p><i>-È dal giorno che ci siamo conosciuti che minacci di dirmi la verità</i> ("it's since the day we met that you've been threatening me to tell me the truth").</p> <p><i>-Al lago cavolo sono tre anni che non ci vado</i> ("to the lake, hell, I haven't gone there for three years").</p>	<p><i>-È in questi giorni che si sta parlando pubblicamente dell'intervento</i> ("it's on these days that people are publicly talking about the intervention")</p> <p><i>-L'ultimo giorno che hanno a disposizione</i> ("the last day that they have at disposal")</p> <p><i>-Un ordine del giorno che ricomprende</i> ("an agenda that includes")</p> <p><i>-I mesi che verranno</i> ("the months that will come")</p> <p><i>-Un ragazzo di 20 anni che ha una bambina di due anni</i> ("a 20-year boy who has a 2-year daughter").</p>
---	--	--	--	--	--

In the phrases above, there are several instances of neo-standard Italian. In line 1 and line 3 (last statements in KIP), anacolutha appear, such as *tua sorella è una che i plumcake ci stravede* ("your sister is one who plum cakes is very fond of"), and *un bel giorno che la fontana non usciva più acqua* ("one fine day that the fountain did not go out any more water"). In these sentences, there is clearly no coherent syntactical connection, and utterances are somehow loose of (standard) grammatical rules.

Other features of neo-standard language are the following ones: the actualising *ci*, as in the statements *è una che i plumcake ci stravede* ("she's one who plum cakes is very fond of them") (see line 1 KIP), and *al lago cavolo sono tre anni che non ci vado* ("to the lake, hell, I haven't gone there for three years") (line 3 OS);

right dislocations (*è una che la sa portare la macchina*, “she’s one who knows how to drive it, the car”, line 1 OS), and duplicated deictics (*quella... là*, “that one... there”, line 3 CHILDES). In the case of the actualising *ci*, is evident that *ci* is redundant and has no spatial or semantic functions. In the first case (*i plumcake ci stravede*), *ci* refers back to plum cakes; in the second, it re-expresses the place (*al lago*) where the person has not been for a while. As regards the right dislocated utterance, the clitic pronoun *la* anticipates the noun phrase *la macchina* (“the car”), functioning as a direct object and appearing in the right periphery of the sentence. In the duplicated deictic utterance (*quella... là*), the spatial adverb *là* reinforces the demonstrative *quella*.

In all examples of line 3, *che* should be replaced by *quando* (“when”) or *in cui* (“in/during which”) in standard Italian. For this reason, the relative clauses introduced by *che* can be considered as substandard. An exception is the last phrase in OS (*al lago cavolo sono tre anni che non ci vado*), where *che* should be replaced by *da quando* (“since when”).

Still in line 3, standard usages appear in PM. For instance, in the phrase *è in questi giorni che si sta parlando pubblicamente* (“it’s on these days that people are publicly talking”), *che* is correctly preceded by the preposition *in* (“on”). In the statement *l’ultimo giorno che hanno a disposizione* (“the last day that they have at disposal”), *che* is a standard conjunction followed by a verb agreeing with “they” (i.e., *che hanno*, with standard subject ellipsis). This means that *che* does not agree back with *giorno*. Another frequent example is *un ordine del giorno che ricomprende* (“an agenda that includes”). In this case, *giorno* pertains to the noun phrase *ordine del giorno* (“agenda”), which assumes a different meaning from the head noun *giorno* (“day”). In addition, *che* is a standard defining relative clause, functioning as a clause subject. For these reasons, the frequency rates of *giorno / sera / mese / anno... che* listed in Table 1 (PM cell) should not be considered as pertaining to neo-standard language.

In the sample phrases of Table 3, standard spoken language features are also noticeable. For example, deixis is found in expressions such as *questo* and *quello* (see CHILDES, KIP and PM in line 1 and line 2), as well as a confirmation token or response elicitor (*vero?*, CHILDES, line 2).

4.3 Samples of cleft sentences

Table 4 reports examples of cleft sentences. There are at least two sample phrases per corpus. Back-translations are written in brackets. Peculiar features of standard and neo-standard spoken language are marked in bold and discussed below.

Table 4. Samples of cleft sentences sourced from the corpora

No.	Corpus / Phraseme	CHILDES	KIP	OS	PM
1	<i>È... che</i>	<p>-È nonna <i>che</i> si mette la lacca? ("is it granny who puts hair spray on?")</p> <p>-È <i>là</i> <i>che</i> piange e cosa succede? ("is it there that s/he cries and what happens?")</p> <p>-Questo è un bambino <i>che gli</i> da un bacio al cane ("this is a child who gives him a kiss to the dog").</p>	<p>-È l'unica casa tua <i>che</i> ho visto forse, eh? ("is it your house only that I saw, uh?")</p> <p>-È meglio <i>che</i> mi fermo da Giacomo? ("is it better that I stay at Giacomo's?").</p>	<p>-È lui <i>che</i> ha problemi ("it's him who has problems")</p> <p>-È un sollievo <i>che</i> non ci siano stati feriti ("it's a relief that there weren't any injuries").</p>	<p>-È anche la prova <i>che</i> l'Europa può cambiare ("it's also the proof that Europe can change")</p> <p>-È da dieci minuti <i>che</i> domando di parlare ("it's for ten minutes that I've been asking to talk").</p>
2	<i>Chi è / com'è / quand'è... che</i>	<p>-Chi è <i>che</i> piange? ("who's the one who's crying?")</p> <p>-Quando è <i>che</i> te lo metti quando mangi? ("when is it that you put it on, when you're eating?").</p>	<p>-Com'è <i>che</i> è tornata poi? ("how is that that she came back, then?")</p> <p>-Chi è <i>che</i> te l'ha regalata? ("who's the one who gave it to you?").</p>	<p>-Com'è possibile <i>che</i> ci sia un cane in un centro commerciale? ("how is it possible that there is a dog in a shopping centre?")</p> <p>-Quand'è <i>che</i> potrò dormire? ("when is it that I can sleep?").</p>	<p>-Com'è <i>che</i> certi servizi vengono riservati ad alcuni? ("how is that that some services are reserved to someone?")</p> <p>-Quand'è <i>che</i> iniziamo a fare un'analisi critica di queste cose? ("when is that that we start analysing these things critically?").</p>

In Table 4, the following traits of neo-standard Italian come to the surface: right dislocation (*questo è un bambino che gli da un bacio al cane*, "this is a child who gives him a kiss to the dog"), (see line 1, CHILDES) and an incorrect use of the subjunctive (*è meglio che mi fermo da Giacomo?*, "is it better that I stay at Giacomo's?") (line 1, KIP).

In the phrase *questo è un bambino che gli da un bacio al cane*, the pronoun *gli* ("to him") redundantly anticipates the indirect object *al cane* ("to the dog") which is placed in the right periphery of the utterance. For this reason, it is a right dislocated

phrase. At the same time, there is an unmarked *da*, which should appear as *dà*. This, however, is due to transcription errors and is not related to the oral production.

Phrases introduced by *è meglio che* must be followed by finites in the subjunctive mode². Therefore, the utterance *è meglio che mi fermo*, should actually be *è meglio che mi fermi*.

In Table 4, there are also instances of standard deictic elements, such as *là* (“there”) (see line 1 CHILDES) and *queste cose* (line 2 PM). The phrase *queste cose* could also be considered as a marker of shared knowledge, or an approximator uttered to reduce the oral production time. In addition, a response elicitor and tag appears in line 1 (KIP) (i.e., *eh?*, “uh?”).

4.4 Samples of presentational *c'è* phrases

Table 5 reports examples of phrases with the presentational *c'è*. There are at least two examples per corpus. Back-translations are written in brackets. Features of standard and neo-standard Italian are marked in bold.

Table 5. Samples of presentational *c'è* phrases sourced from the corpora

No.	Corpus / Phraseme	CHILDES	KIP	OS	PM
1	<i>C'è... che</i>	<p><i>-C'è il cane che gli corrono dietro le api</i> (“there’s the dog that the bees run after it”)</p> <p><i>-C'è una donna che gli da un libro a una bambina</i> (“there’s a woman who gives him a book to a girl”).</p>	<p><i>-Non c'è bisogno che lo fai tu</i> (“there’s no need for you to do it”)</p> <p><i>-C'è quella matrimoniale che ha il bagno</i> (“there’s the master bedroom that has a bathroom”)</p> <p><i>-C'era (...) la vecchietta che gli squilla il telefono</i> (“there was... the oldie / old lady who the telephone rings”).</p>	<p><i>-C'è una cosa che ho imparato</i> (“there’s a thing that I learnt”)</p> <p><i>-C'è uno stalker che la perseguita</i> (“there’s a stalker who harasses her”).</p>	<p><i>-C'è qualche senatore che vota per la prima volta</i> (“there’s some senator who votes for the first time”)</p> <p><i>-C'è un altro fenomeno che avevamo contestato</i> (“there’s also another phenomenon that we contested”).</p>

² See <https://aulalingue.scuola.zanichelli.it/benvenuti/2008/11/11/il-congiuntivo/>.

2	<i>Vi è... che</i>	-	-	<i>-Vi è uno che non sembra d'accordo</i> ("there's a someone that seems to disagree") <i>-Non vi è dubbio che Dick ha un talento speciale</i> ("there's no doubt that Dick has a special talent").	<i>-Vi è una seconda questione che francamente mi ha sorpreso</i> ("there's a second issue that has frankly surprised me") <i>-Non v'è dubbio che non vi siano obblighi personali</i> ("there's no doubt that there aren't any personal obligations").
---	--------------------	---	---	--	---

In line 1, there are several traits of neo-standard language, such as left and right dislocations, substandard relative clauses and unmarked usages of *gli* ("to him"). In particular, in line 1 (CHILDES), *c'è il cane che gli corrono dietro le api* ("there's the dog that the bees run after it") is a left-dislocated utterance characterised by a substandard relative clause. The pronoun *gli* follows and re-states the subject *il cane* ("the dog"), which is placed in the left periphery of the sentence. For this reason, it is a left-dislocated statement. In addition, *che* ("that") should actually be *a cui* or *al quale* ("after which/whom"). Still in line 1 (CHILDES), there's right dislocation and an unmarked use of the *gli* pronoun in the utterance *c'è una donna che gli da un libro a una bambina* ("there's a woman who gives him a book to a girl"). This statement is dislocated to the right as the pronoun *gli* ("to him") (which should be *le*, "to her") precedes the head noun *a una bambina*, functioning as an indirect object ("to a girl") and appearing in the right periphery of the sentence. In addition, *da* should be transcribed as *dà*.

In line 1 (KIP) (in the utterance *c'era (...) la vecchietta che gli squilla il telefono*, "there was... the oldie / old lady who the telephone rings"), there are other neo-standard elements. In particular, these are as follows: an unmarked use of *gli*; a substandard relative clause, and an incorrect use of the present tense. Firstly, *gli* should be *le*, as the person referred to is a woman (*vecchietta*, "old lady"); secondly, *che* should be changed into *a cui* or *alla quale* ("to/for whom"); thirdly and lastly, the finite *squilla* ("rings") should be *squillava* ("rang"), as the statement starts with *c'era* ("there was").

In line 2 (OS), only an unmarked use of the indicative mode appears. In the statement *non vi è dubbio che Dick ha un talento speciale* ("there's no doubt that Dick has a special talent"), *ha* should be replaced by *abbia*, as phrases starting with *non vi è dubbio che* must be followed by a finite in the subjunctive mode³. As a matter of fact, line 2 (PM) reads *non v'è dubbio che non vi siano obblighi*

³ See: https://www.treccani.it/magazine/lingua_italiana/domande_e_risposte/grammatica/grammatica_1315.html.

personali (“there’s no doubt that there aren’t any personal obligations”), where the finite *siano* is subjunctive.

As far as standard features of spoken Italian are concerned, line 1 (KIP) shows *quella* (“that one”), which is a deictic element.

5. Discussion

This paper highlighted marked differences in the frequencies of some neo-standard expressions such as *è uno/una che* and *è quello/quella che* (particularly characterising spontaneous adult conversations); *chi è che*, *com’è che* and *quand’è che* (hallmarking children’s conversations), and *è... che* and *vi è che* (mostly appearing in parliamentary debates). Scripted dialogues such as film subtitles did not yield any relevant results or particularly high occurrences. This may be due to the fact that the subtitles analysed were translations from dialogues in other languages, rather than transcripts of film dialogues in Italian. If this is the case, this fact confirms scholarly research regarding the reluctance of film language adaptors to use traits of neo-standard Italian (Pavesi and Perego 2006).

As a whole, multifunctional *che* phrases were retrieved across all corpora. The only expressions which were used in a standard manner were *giorno / sera / mese / anno... che* in PM. Cleft sentences were also found in all the corpora investigated, as well as presentational *c’è* constructions. In the latter case, however, phrasemes with *vi è... che* or *v’è... che* appeared only in PM and OS.

With regard to other neo-standard peculiarities, the analysis provided some additional insightful traits, such as the use of the actualising *ci* (in KIP and OS); duplicated deictics (in CHILDES); substandard relative clauses (across all corpora); left and right dislocations (in CHILDES and KIP); erroneous usages of *gli* instead of *le* (in CHILDES and KIP); erroneous usages of the subjunctive mode (in KIP and OS) and of tenses (in KIP), as well as anacolutha (in KIP). It is evident that such results are only related to the phrases extracted from the corpora. Therefore, more or other neo-standard traits could appear across all corpora, if other phrases or phrasemes were analysed.

In addition to neo-standard features, the analysis brought to the surface standard traits of spoken language, such as deixis, response elicitors and tags.

6. Conclusions

This paper was aimed at exploring the features of neo-standard Italian in different spoken circumstances and settings. To this aim, four different corpora were focused on: children’s conversations (CHILDES), adult conversations (KIParla), parliamentary debates (ParlaMint), and film subtitles (Open Subtitles – Italian). In particular, a small-scale quantitative and qualitative comparative analysis was carried out. The quantitative aspect was made explicit via the relative frequencies across the corpora, whereas the qualitative analysis was developed by extracting phrases or statements with the phrasemes focused on.

The findings show that the traits addressed, i.e., multifunctional *che*, cleft sentences and presentational *c'è* were noticed in the statements or phrases extracted from the four corpora. This means that, irrespective of the context or setting (i.e., less formal as in children's conversations vs. more formal such as in parliamentary speeches), the above traits appeared in all the corpora investigated. To some extent, certain phrasemes characterised some corpora more prevalently. For example, *chi è che*, *com'è che* and *quand'è che* strongly hallmarked children's conversations, whereas *è... che* and *vi è... che* mostly appeared in parliamentary debates. Also, anacolutha featured adult spontaneous conversations, as well as incorrect usages of modes or tenses. These results, however, relate only to the phrasemes dealt with and, most of all, they refer to the statements or utterances extracted from each corpus. Therefore, additional neo-standard traits could come to the fore in some or in all corpora, if more extended or in-depth analyses were carried out.

In addition to neo-standard features, the investigation brought to the fore standard traits of spoken language, such as deixis and response elicitors.

The research question posed in this paper asked whether there are any differences in the frequencies of usage and in the types of traits of neo-standard Italian in the spoken corpora investigated. This question has already been addressed, and the answer is affirmative. This paper showed not only different relative frequencies of occurrence across the corpora, but also some nuanced differences in the types of phrasemes used. For example, expressions starting with *è... che* mostly appear in parliamentary speeches, as well as *vi è... che* phrases. Conversely, phrasemes with *giorno / sera / mese / anno... che* are used in a standard manner in parliamentary speeches. Hence, the related high frequencies do not point to neo-standard language.

The limits of this paper lie in the fact that only a few statements per corpus were extracted and analysed. If more utterances could have been retrieved, the findings might have been more comprehensive. Nonetheless, the search methodology rigorously shows the corpus queries and syntax (see Table 2). Therefore, the replication of the analysis is possible and further investigations could encompass other traits of neo-standard Italian in the same or in other spoken corpora. In addition, more utterances could be investigated per corpus, thus allowing the generalisation of the findings.

References

- Alfonzetti, Giovanna. 2002. *La Relativa Non-Standard. Italiano Popolare o Italiano Parlato?* Palermo: Centro Studi Filologici e Linguistici Siciliani.
- Ballarè, Silvia. 2020. L'italiano Neo-Standard Oggi: Stato Dell'Arte. *Italiano LinguaDue*, 20, 469-492.
- Berruto Gaetano. 1985. Per una Caratterizzazione del Parlato: L'Italiano Parlato Ha un'Altra Grammatica?. In Günter Holtus and Edgar Radtke (eds.), *Gesprochenes Italienisch in Geschichte und Gegenwart*, 120-153. Tübingen: Gunter Narr Verlag.
- Berruto, Gaetano. 1987. *Sociolinguistica dell'Italiano Contemporaneo*. Rome: La Nuova Italia Scientifica.

- Berruto, Gaetano. 2017. What Is Changing in Italian Today? In Massimo Cerruti, Claudia Crocco and Stefania Marzo (eds.), *Towards A New Standard: Theoretical and Empirical Studies on the Restandardization of Italian*, 31-60. Berlin / Boston: De Gruyter Mouton.
<https://doi.org/10.1515/9781614518839-002>
- Chekalil, Aicha. 2020. La Dislocazione a Sinistra nella Didattica dell'Italiano LS a Studenti Adulti: Premesse Teoriche e Proposte Didattiche. *Bollettino Itals*, 18(83), 13-38.
- D'Achille, Paolo. 2019. *L'Italiano Contemporaneo*. Bologna: Il Mulino.
- Erjavec, Tomaz, Ogrodniczuk, Maciej, Osenova, Petya, Ljubešić, Nikola, Simov, Kiril, Grigorova, Vladislava, Rudolf, Michał, Pančur, Andrej, Kopp, Matyáš, Barkarson, Starkaður, Steingrímsson, Steinþór, van der Pol, Henk, Rayson, Paul Kopp. 2021. *Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 2.1*, Slovenian language resource repository CLARIN.SI. <http://hdl.handle.net/11356/1431>.
- Gavazzi, Stefania. 2009. La Dislocazione a Destra nella Didattica dell'Italiano L2/LS: Riflessioni e Proposte Didattiche. *ElLe*.
<https://www.itals.it/la-dislocazione-destra-nella-didattica-dell%e2%80%99italiano-l2ls-riflessioni-e-proposte-didattiche>.
- MacWhinney, Brian, and Snow, Catherine. 1985. The child language data exchange system. *Journal of Child Language*, 12(2), 271-295. <https://doi.org/10.1017/S0305000900006449>
- Mauri, Caterina, Silvia Ballarè, Eugenio Goria, Massimo Cerruti & Francesco Suriano. 2019. KIParla corpus: a new resource for spoken Italian. In Raffaella Bernardi, Roberto Navigli, and Giovanni Semeraro (eds.), *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it*, Bari, Italy, November 13-15. <https://ceur-ws.org/Vol-2481/>.
- Menin, Roberto. 2014. The Concept of Performability and Its Application within Theatrical Tradition. *Studi Germanici*, 5, 109-137. <https://www.studigermanici.it/studi-germanici-5-2014/>.
- Pavesi, Maria, and Perego, Elisa. 2006. Profiling Film Translators in Italy: a Preliminary Analysis. *The Journal of Specialised Translation*, 6, 99-114. <https://doi.org/10.26034/cm.jostrans.2006.741>
- Poletto, Cecilia, and Sanfelici, Emanuela. 2017. Relative Clauses. In Andreas Dufter and Elisabeth Stark (eds.), *Manual of Romance Morphosyntax and Syntax*, 804-836. Berlin / Boston: De Gruyter. <https://doi.org/10.1515/9783110377088-022>.
- Sabatini, Francesco. 1985. 'L'Italiano dell'uso Medio': Una Realtà tra le Varietà Linguistiche Italiane. In Günter Holtus and Edgar Radtke (eds.), *Gesprochenes Italienisch in Geschichte und Gegenwart*, 154-184. Tübingen: Gunter Narr Verlag.
- Tiedemann, Jörg. 2012. Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'2012)*, 2214-2218. http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf.
- Voghera, Miriam. 1992. *Sintassi e Intonazione nell'Italiano Parlato*. Bologna: Il Mulino.

Patrizia Giampieri, MSc, PhD, is an Associate Professor of English at the University of Perugia (Italy). She is a member of the Executive Board of EST (European Society for Translation Studies). She has authored books and academic papers on English as a second language, legal English, corpus-based translation, machine translation, AI-driven translation, and spoken language.