Silvia Ballarè and Guglielmo Inglese (Eds.)
Sociolinguistic and Typological Perspectives on Language Variation

Trends in Linguistics Studies and Monographs

Volume 374

Sociolinguistic and Typological Perspectives on Language Variation

Edited by
Silvia Ballarè and Guglielmo Inglese

DE GRUYTER
MOUTON

ISBN 978-3-11-078106-9
e-ISBN (PDF) 978-3-11-078116-8
e-ISBN (EPUB) 978-3-11-078123-6
ISSN 1861-4302

Library of Congress Control Number: 2023940160

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.dnb.de.

Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Contents

Caterina Mauri, Alessandra Barotto and Simone Mattiola

6 Counterfactual conditionals: Linguistic variation in Italian and beyond

Abstract: The aim of this paper is to integrate the analysis of the intra-linguistic variation attested in the coding of counterfactual conditionals in spoken Italian with a larger cross-linguistic perspective. We start by providing a sociolinguistically-informed discussion of the counterfactual conditional strategies attested in spoken Italian, through a corpus-based analysis. After describing the corpus and the parameters of analysis, we provide a typology of the attested constructions, based on the verbal forms employed for protasis and apodosis, showing that the observed variation goes beyond the representations available in the literature. In particular, we describe the spread of counterfactual imperfective past indicative forms, which are especially frequent in symmetrical constructions and are argued to be associated to lower educational achievements and, to a lesser extent, to informal contexts. We then consider the picture of Italian counterfactual conditionals within a wider perspective, to verify whether the patterns described for Italian follow more general tendencies. After addressing some crucial methodological issues, concerning the problems raised by the integration of a sociolinguistic and a cross-linguistic approach, we argue that Italian data are probably to be analyzed as manifestations of a widespread trend towards the use of symmetrical counterfactual constructions on the one hand and the use of past habituals for functions connected to the irrealis domain on the other hand.

1 Introduction

1.1 Counterfactual conditionals: Defining the object of analysis

In the literature, different types of conditional constructions are usually identified on the basis of their factuality status (cf. Zaefferer 1991; Sweetser 1990; Taylor 1997; Thompson et al. 2007), even though the number of types and the terminology tend to vary. Thompson et al. (2007), following Schachter (1971), propose a two-way clas-

Caterina Mauri, University of Bologna, e-mail: caterina.mauri@unibo.it
Alessandra Barotto, University of Insubria, e-mail: alessandra.barotto@uninsubria.it
Simone Mattiola, University of Bologna, e-mail: simone.mattiola@unibo.it

https://doi.org/10.1515/9783110781168-006

sification, distinguishing between *reality* conditionals and *unreality* conditionals. Reality conditionals refer to 'real' present, 'habitual/generic', or past situations, whereas unreality conditionals refer to situations in which we imagine what might be or what might have been (*imaginative* conditionals) or situations in which we predict what will be (*predictive* conditionals). Imaginative conditionals are further subdivided into *hypothetical* (what might be) and *counterfactual* (what might have been) conditionals.

In this paper, we will focus on the latter, thus considering constructions conveying a conditional relation between two events that the speaker acknowledges as not realized.[1] If on the one hand counterfactual conditionals are associated to the domain of so-called irrealis (cf. Elliot 2000; Mauri and Sansò 2012), on the other hand within conditional types they happen to be the most connected to an epistemic condition of certainty, given the fact that both protasis and apodosis are known to have *not* occurred. In all the other types, indeed, the occurrence of the conditional relation is a possibility, though with different probability expectations, and the speaker is necessarily in an epistemic state of uncertainty.

Taylor (1997: 301–302) opposes *counterfactual* conditionals to *factual* conditionals on the one hand and to *hypothetical* conditionals on the other hand (emphasis and italics added):

> In a **factual** conditional, the content of the if-clause is *presumed* to be the case, whilst in a **counterfactual** the content of the if-clause is *taken to be* contrary to fact. Between these categories stand the **hypothetical** conditionals, in which the content of the if-clause is entertained as a *possibility*, neither in accordance with reality, nor necessarily inconsistent with it.

Taylor (1997: 302) notes that is possible to propose a "gradience of epistemic likelihood of the protasis", ranging from factual conditionals, in which the content of the *if*-clause is presumed to be the case at some point as in (1a), to hypothetical conditionals, in which the content of the *if*-clause is entertained as a possibility as shown in (1b), to counterfactuals, in which the content of the *if*-clause is taken to be contrary to fact as shows in (1c).

(1) a. *If he said that (and we heard him), he's a liar.* (factual)
b. *If he said that (and we don't know if he did), he'd be a liar*. (hypothetical)
c. *If he had said that (and we know that he didn't), he would be a liar.* (counterfactual)

1 This paper is the result of a continuous collaboration between the authors, who jointly wrote §1. Caterina Mauri is responsible for §2, Alessandra Barotto is responsible for §3.2 and §3.3, Simone Mattiola is responsible for §3.1 and §4.

This notion of "gradience of epistemic likelihood" can be found also in Comrie (1986), who in turn proposes a different approach that rejects classifications based on discrete categories. Comrie suggests that conditional types fall along a continuum of hypotheticality, with reality conditionals at the higher end and counterfactual conditionals at the extreme lower end. This continuum can thus be segmented differently by natural languages, which can express linguistically different distinctions.

Despite the differences that can be found in the literature, counterfactual conditionals are generally considered the type with the lower degree of factuality or likelihood of actualization. Indeed, by definition, counterfactual conditionals represent an interesting conundrum for philosophy and logic, since they "hypothetically present[s] as having happened an event which did not happen" (Lazard 2006: 61).

Interesting tendencies have been identified with regard to counterfactual conditionals in sociolinguistic and typological approaches. For instance, Singaporean English has been observed to have variable marking for counterfactuality (cf. Crewe 1984). Ziegeler (1994) notes that in Singaporean English counterfactual ideas and concepts are encoded through different means (e.g., *wish*, *suppose/supposing*) and "are not systematically similar to the means of expressing distinctions of temporality and modality in standard varieties" (Ziegeler 1994: 46). Haiman and Kuteva (2002) identify a cross-linguistic tendency for the use of symmetric counterfactual constructions (i.e. having the same verbal forms in protasis and apodosis), especially in nonstandard varieties: according to them, in counterfactuals it is frequent that "the protasis and the apodosis clauses, irrespective of their particular morphological form, have parallel structures" (Haiman and Kuteva 2002:102). They provide examples of nonstandard English, French, Spanish, Bulgarian, Welsh and propose an explanation in terms of iconicity. A recent study dedicated to counterfactuals from a typological perspective is the one by Olguín Martínez and Lester (2021): based on a 106-language sample, they conclude that the symmetry of counterfactuals is dependent on the irrealis *vs*. realis (or unmarked) coding of the protasis. They argue that "symmetrical systems tend to be those that treat the entire counterfactual conditional construction as ungrounded or hypothetical. Asymmetrical systems tend to be those which afford special status (either grounded or unmarked) to the protasis while leaving the apodosis non-actualized" (Olguín Martínez and Lester 2021: 175).

The case of Italian is challenging with respect to the existing literature. In standard Italian, counterfactual conditionals are asymmetric and require the use of a subjunctive form in the protasis and conditional form in the apodosis ((2a), cf. Wiberg 2010). However, as shown in example (2b), another construction is well attested especially in spoken language, in which the past imperfective indicative form is used both in the protasis and in the apodosis. This strategy has been classi-

fied as substandard in the literature (see Berruto 1983: 59; Bertinetto 1986; Berretta 1993; Mazzoleni 2013) and appears to be attested mainly for counterfactuals but marginally also for hypothetical conditionals.

(2) Italian

a. *Se lo avessi spedito stamattina, sarebbe arrivato domani.*
if it AUX:SUBJ.PAST:1SG send:PAST.PTCP this.morning AUX:COND:3SG arrive:PAST.PTCP:M tomorrow
'If you had shipped it this morning, it would have arrived tomorrow.' (standard Italian)

b. *Se lo spedivi stamattina arrivava domani*
if it send:IND.IMPF.PAST:2SG this.morning arrive:IND.IMPF.PAST:3SG tomorrow
'If you had shipped it this morning, it would have arrived tomorrow.' (Mazzoleni 2013: 5)

Apparently, then, Italian shows a standard asymmetric construction (2a) and a nonstandard symmetric construction (2b), in line with Haiman and Kuteva's (2002) observations, but in the latter it involves the use of past imperfective indicative forms, which would be classified by Olguín Martínez and Lester as actualized, thus going against their prediction.

1.2 Aims and overview

The aim of this paper is to provide a corpus-based, sociolinguistically-informed analysis of the intra-linguistic variation attested in the coding of counterfactual conditionals in spoken Italian, focusing on the construction symmetry and the verbal forms employed in protasis and apodosis (§2). The results will be then compared to the wider picture of cross-linguistic variation, in order to understand and explain the features observed in Italian also in the light of typological patterns (§3).

After describing the corpus and the parameters based on which Italian data have been examined (§2.1), in §2.2 we provide a typology of the attested constructions, based on the verbal forms employed for protasis and apodosis, showing that the observed variation goes beyond the binary representations available in the literature (cf. (2)). In particular, we will focus on the spread of counterfactual imperfective past indicative forms, which are especially frequent in symmetrical

constructions (as exemplified in (2b)) and will be argued to be associated to lower educational achievements and to informal contexts (§2.3).

In §3 we aim to analyze the rise and spread of symmetric counterfactual constructions in the light of the attested cross-linguistic variation. In §3.1 we address some crucial methodological issues, concerning some problems raised by the integration of the intra-linguistic and inter-linguistic perspectives. A preliminary typological analysis is then described in §3.2, showing that the attested cross-linguistic variation confirms both the widespread use of symmetrical counterfactual constructions, if compared to other types of coniditonals, and the crucial role played by past forms. A focus on the relation between past habituals and irrealis is finally presented in §3.3, with the aim to explain why the symmetric construction of Italian shows an allegedly actualized form, such as the indicative past imperfective, going against Olguín Martínez and Lester's (2021) predictions. Finally, some conclusive remarks will be presented in §4.

2 Counterfactual conditionals: A sociolinguistic approach to spoken Italian

When it comes to counterfactual conditional constructions in spoken Italian, two main strategies have been acknowledged in the literature (Berruto 1983; Bertinetto 1986; Berretta 1993): the strategy that is ascribed to the standard variety includes the use of a past subjunctive form in the protasis and a past conditional in the apodosis (as exemplified in (2a) and (3), henceforth SUB-COND strategy), while the strategy that is ascribed to the nonstandard varieties (substandard, according to Berruto 1983) includes the use of a past imperfective indicative form both in the protasis and in the apodosis (as exemplified in (4) and (5), henceforth DOUBLE-IND strategy).

(3) ParlaTO, PTA18

01	TOI031	*ovviamente*	***se***	***poi***	***avesse***		***partorito***	
		obviously	if	then	AUX:SUBJ.PAST:3SG		give.birth:PAST.PTCP	
		'obviously if then she had'						
02		***ci***	***saremmo***		***dovuti***		***fermare***	*in*
		REFL.1PL	AUX:COND:1PL		must:PAST.PTCP:M.PL		stop:INF	in
		ambulanza						
		ambulance						
		'we should have stopped in an ambulance'						

(4) KIP, TOA3001

01 TO029 *immaginate* *così x* *una* *scena* *del* *genere*
imagine:IMP.2PL so INDEF scene of.DEF genre
'Imagine this, a scene like this'

02 TO033 *certo* *sì* *che* *non* *hai* *neanche* *tempo* *x x*
sure yes that NEG have:IND.PRES.2SG even time X X
cioè *hai* *capito*
I.mean you know
'sure yes, that you don't even have time, I mean, you know'

03 TO029 *ah* *la* *rossi* *ci* ***faceva il culo*** *soltanto* *se* *lo*
ah DEF Rossi 1PL.DAT scold:IND.IMPF.PAST:3SG only if it
pensavamo *di* *fare* *una* *roba* *del* *genere*
think:IND.IMPF.PAST:3SG of do:INF INDEF thing of:DEF genre
'ah, the Rossi would have scolded us if we had even thought about doing such a thing'

(5) ParlaTO, PTD006
se *non* ***trovavo*** *niente* *m'* ***attaccavo***
if NEG find:IND.IMPF.PAST:1SG nothing 1SG.ACC hang:IND.IMPF.PAST:1SG
al *tram*
to:DEF tram
'if I had not found anything I could have sung for it'

Both in the SUB-COND and in the DOUBLE-IND construction the verb forms are in the past, but what crucially differentiates the two strategies is the mood. While conditional and subjunctive are typically associated to irrealis and potential states of affairs (cf. Mauri and Sansò 2016), indicative is the mood for realis and actualized events. Furthermore, whereas in the Italian subjunctive and conditional paradigms there is no alternative between perfective and imperfective past forms,[2] in the indicative mood speakers may choose between two perfective past forms, namely simple past (*passato prossimo*) and perfective remote past (*passato remoto*), and one imperfective past form (*imperfetto*),[3] depending on aspectual features and on

2 In the subjunctive paradigm, the choice between so-called 'past subjunctive' (*che abbia partorito*), 'imperfect subjunctive' (*che partorisse*) and 'pluperfect subjunctive' (*che avesse partorito*) in subordinate clauses is governed by the tense selected in the main clause (cf. Bertinetto 1986).

3 There are also two additional past forms, the first composed by the auxiliary inflected for the remote past followed by a past participle (*trapassato remoto*) and the second composed by the auxiliary inflected for the imperfective past followed by a past participle (*trapassato prossimo*). They are used to refer to some anteriority relation with respect to a past event.

how recent is the past event they are referring to. The indicative forms attested in counterfactual strategies show a specific type of past, namely imperfective.

Indicative imperfective past (henceforth PastImpf indicative) is employed in spoken Italian mainly for descriptions in the past, to refer to states of affairs taking place at the same (past) time, and to denote habitual states or events in the past. In addition to these three uses, it is the only indicative past form that also shows functions related to the domain of imagination, possibility or courtesy, where subjunctive or conditional would be required. We find it for example in fictitious contexts, as in (6) where it is introduced by *come se* 'as if', or in children's fantasy games.

(6) ParlaTO, PTB007

01	TOI021	*no*	*perché*	*tanto*	*s(e)*	*andavi*		*a*	
		no	because	anyway	if	go: IND.IMPF.PAST:2SG		to	
		chiederglieli							
		ask:3SG.DAT:3PL.OBJ.M.PL							
		'no, because anyway if you went to ask them [the money] to them'							
02		*è è è*		***come***	***se***	***tu***	***dovevi***		
		it.is it.is it.is		as	if	you	must:IND.IMPF.PAST:2SG		
		'it's as if you had to'							
03		*dare*	*tu*	*dei*	*soldi*	*a*	*loro*	*e*	*non*	*che*	*lo(ro)*
		give	you	INDEF.PL	money	to	them	and	NEG	that	they
		'it was you who had to give money to them and not them who'									
04		***dovevano***		*darli*		*a*	*te*				
		must:IND.IMPF.PAST:3PL		give:3PL.OBJ.M		to	you				
		'had to give money to you'									

PastImpf indicative is also widespread in polite requests, instead of the conditional: in (7) we note, at the beginning of an interview, the expression *volevo sapere* 'I wanted to know', which clearly does not refer to an earlier stage, but replaces the conditional *vorrei sapere* 'I would like to know'. These uses derive from an anaphoric reference to some previous context motivating the speech act, which thus appears to be justified.

(7) KIP, BOD2016

01	BO119	*okay*	
		okay	
		'okay'	
02		*ci*	*siamo*
		LOC	be:IND.PRES.1PL
		'we're here'	

03 *allora ciao*
then hello
'hello then'

04 BO120 *ciao*
hello
'hello'

05 BO119 *tu ti chiami*
you REFL.2SG call:IND.PRES.2SG
'your name is'

06 BO120 *francesco*
Francesco
'Francesco'

07 BO119 *okay franci allora ehm*
okay Franci so ehm
'okay Franci, so ehm'

08 ***volevo*** ***sapere*** *eh tu vivi qua*
want:IND.IMPF.PAST:3PL know eh you live:IND.PRES.2SG here
a bologna
in Bologna
'I wanted to know eh you live here in Bologna'

09 BO120 *sì*
yes
'yes'

The use of PastImpf indicative in counterfactual conditionals (as in (4) and (5)) has been analyzed as a typical feature of nonstandard varieties, not only for Italian (Berruto 1983; Bertinetto 1986; Berretta 1993), but also for other languages (cf. Ziegeler 1994 for Singapore English). However, little if any corpus evidence has been provided in the literature.

In order to fill this gap, in this section we will provide a systematic analysis of how counterfactual conditionals are conveyed in spoken Italian, based on the KIParla corpus (Mauri et al. 2019). Since this resource allows to access metadata regarding both speakers and conversations, we will connect the variation observed in the data with two main sociolinguistic variables, namely (i) conversation formality on the one hand, and (ii) speakers' educational achievement on the other hand. We aim to understand whether the strategies attested in spoken Italian to express counterfactual conditionals have a non-random distribution across the diaphasic and diastratic axes, thus showing preferential associations with specific registers and/or speakers with particular educational achievements.

2.1 Data and parameters of analysis

The KIParla corpus of spoken Italian (Mauri et al. 2019) is a resource characterized by a modular and incremental structure. At present (July 2022), it consists of two modules, which together amount to 1.125.996 words (ca. 110 hours of recording).[4] For both modules, it is possible to search transcriptions aligned with the audio files, accessing the metadata regarding both speakers and conversations.

The KIP module consists of 70 hours of recorded speech collected in Turin and Bologna between 2016 and 2019. It has relatively few internal diastratic variation, if we consider the educational degree, because it only comprises conversations by speakers with higher educational achievement (university students and professors). On the contrary, it shows a good range of diaphasic variation, including both informal interaction (free conversation, semi-structured interviews) and formal interaction (lessons, exams, office hours). The ParlaTO module consists of 40 hours of recorded speech collected in Turin between 2018 and 2019. It has no internal diaphasic variation because it includes only informal interaction (semi-structured interviews). It shows instead a good degree of diastratic variation, including speakers with varying levels of educational achievement and varying employment. The joint analysis of the modules allows to integrate both diastratic and diaphasic variation in the study.

For the aim of this research, we extracted and annotated all the conditional constructions introduced by *se* 'if' attested in the KIP and in ParlaTO modules (tot. 1856 occurrences), out of which we identified 104 occurrences of counterfactual conditional constructions that constitute our dataset. We considered both linguistic and extra-linguistic parameters, which will be employed as predictors in §2.3. The list of parameters is summarized in Table 1:

Table 1: Linguistic and extra-linguistic parameters.

Linguistic parameters		**Extra-linguistic parameters**	
(i)	Verb form for protasis and apodosis	(v)	Formality of the interaction
(ii)	Symmetry of the construction	(vi)	Speakers' educational achievement
(iii)	Order of protasis and apodosis	(vii)	Speakers' age
(iv)	Subject agreement on the verbs	(viii)	Speakers' origin

As far as the linguistic predictors are concerned, we annotated the (i) verb form used for protasis and apodosis (past conditional, past subjunctive or PastImpf indicative),

4 The KIParla Corpus is freely available at www.kiparla.it

(ii) the symmetry *vs.* asymmetry of the construction (same verbal form or different verb form for protasis and apodosis), (iii) the respective order of protasis and apodosis (preposed *vs.* post-posed protasis) and (iv) the subject agreement of the two verb forms (1s/p, 2s/p, 3s/p, impersonal, for both verbs in protasis and apodosis).

As for the extra-linguistic predictors, we considered the (v) formality of the interaction (formal *vs.* informal), (vi) the educational achievement of the speaker (higher *vs.* lower), (vii) age, (viii) origin. Formality and educational achievement have been considered as binary parameters (Table 2).

Table 2: Interaction types and formality; degree type and educational achievement.

FORMALITY OF THE INTERACTION	
Formal interaction	**Informal interaction**
– lessons	– free conversations
– exams	– semi-structured interviews
– office hours	
LEVEL OF EDUCATIONAL ACHIEVEMENT	
Lower educational achievement	**Higher educational achievement**
– elementary degree	– secondary school degree and attending university
– middle school degree	– university degree
– secondary school degree and working	– phd

Lectures, exams and office hours were classified as formal interactions, while free conversation and semi-structured interviews were classified as informal interactions. Speakers who studied until elementary or middle school were grouped with working graduates, while university students were assimilated to graduates and PhDs.

2.2 Counterfactual constructions in Spoken Italian

If we consider the 104 occurrences of counterfactual constructions attested in ca. 110 hours of spontaneous speech, we first of all observe a greater variation of constructions than the SUB-COND *vs.* DOUBLE-IMPF opposition, exemplified in (3) and (4)-(5), respectively. As shown by Figure 1, the category OTHER indeed corresponds to roughly one fourth of the total (23%) and in the ParlaTO module it reaches the 27% of occurrences, exceeding the percentage of SUB-COND strategies (see Figure 3). Let us first of all see what types of strategies fall within the label OTHER.

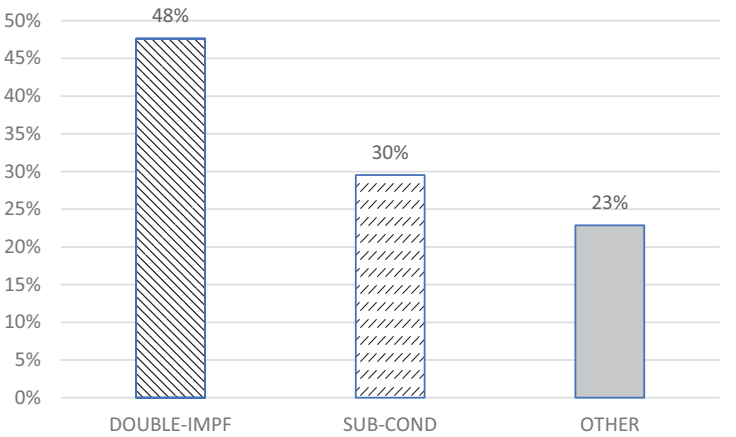

Figure 1: Counterfactual strategies in the KIParla Corpus.

Under the label OTHER we find several combinations of verbal forms, each of which has a low occurrence rate. Yet, the relatively high ratio of strategies deviating from what is expected is revealing of the high degree of fluctuation attested in the expression of counterfactual conditionals, which are rare enough to leave room for a certain indecision by speakers, who frequently find themselves self-repairing in the search for the 'correct' verb form to use. We can observe an example of self-repair in (9) (line 03): the speaker starts to utter the standard conditional form in the apodosis (*ci sar. . .*) but then stops and self-repairs (*cioè* 'I mean') by choosing the PastImpf indicative *potevo*. In this case, the speaker is uncertain in using the standard strategy and decides to recur the indicative.

Under OTHER we find the following combinations (a-c):[5]

a) PastImpf indicative in the protasis + past conditional in the apodosis, exemplified in (8):

(8) ParlaTO, PTA001
01 TOI001 *eh beh tutto è nato da bambino quando mia madre mi disse un giorno che*
'everything started when I was a child, when my mother once told me that'

5 In these examples, glosses are provided only for the relevant verb forms.

02 ***se*** ***tifavo*** *juve* *mio* *padre* *non*
if support:IND.IMPF.PAST:1SG Juventus my father NEG
sarebbe
AUX:COND.3SG
'if I had supported Juventus, my father would'
03 ***tornato*** *più* *a* *casa*
come.back. PAST.PTCP:M.SG anymore to home
'never have returned home'

b) Past subjunctive in the protasis + PastImpf indicative in the apodosis, exemplified in (9):

(9) KIP, BOD2008
01 BO053 [. . .] *magari* *se* *qualcuno* *m'* ***avesse*** ***un***
maybe if someone 1SG.ACC AUX:SUBJ.PAST:3SG a
minimo ***spinto***
bit push:PAST.PTCP
'[. . .] maybe if someone had pushed me a bit'
02 *e* *m'* ***avesse*** ***detto*** *dai* *oh*
and 1SG.ACC AUX:SUBJ.PAST:3SG say:PAST.PTCP come.on oh
andiamo
let's.go
'and had told me *come on oh let's go*'
03 *probabilmente* *ci* *sar* *cioè* ***potevo***
probably LOC AUX:COND.1SG I.mean can:IND.IMPF.PAST:1SG
anche *andarci*
also go.there
'probably I would. . . I mean I could have gone'

c) Indicative present or simple past in the protasis + PastImpf indicative in the apodosis, exemplified in (10):

(10) ParlaTO, PTB014
Se ***vado*** *dritto* ***andavo*** *a* *san peyre*
if go:IND.PRES:1SG straight go IND.IMPF.PAST:1SG to San Peyre
'if I had gone straight I would have arrived to San Payre'

The only symmetric strategy attested in the KIParla is the DOUBLE-IMPF construction exemplified in (4) and (5), while SUB-COND and OTHER strategies are all asymmetric, since they show different verbal forms for protasis and apodosis. Furthermore,

the great majority (87%) of the counterfactual constructions attested in our dataset show the protasis preceding the apodosis. If we consider the relative frequencies of person agreements on the two verb forms (Figure 2), we see that 3rd person singular is the most frequent one in both protasis and apodosis, immediately followed by 1st person singular and 3rd person plural. The use of impersonal forms is more frequent in the apodosis, while 2nd person singular is easier to find in the protasis (in some cases, it is a generic 2nd person, not referring to the actual interlocutor).

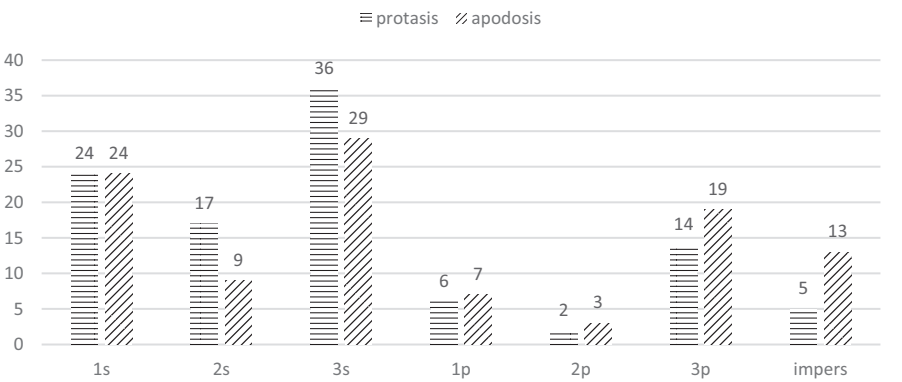

Figure 2: Subject agreement on verb forms in protasis and apodosis (in this chart, the absolute numbers are provided).

Going back to the general picture represented in Figure 1, we can see that the DOUBLE-IMPF strategy is by far the most frequently attested in the KIParla corpus, covering almost half of the total occurrences, while the SUB-COND strategy covers one third of the total. Given the different socio-demographic composition of the two KIParla modules, it is worth considering the distribution of the attested counterfactual constructions separately in the KIP and ParlaTO corpora.

Figure 3 indeed shows that there is a significant difference between the two corpora, especially concerning the respective ratio of the SUB-COND and DOUBLE-IMPF strategies: in the KIP module the SUB-COND construction is the most frequently used one (47%), immediately followed by the DOUBLE-IMPF one (40%), while in the ParlaTO module we observe a much higher percentage for the DOUBLE-IMPF strategy, attested in the 57% of the cases, and an extremely lower percentage for the SUB-COND strategy, which is attested only the 16% of the occurrences, and is exceeded by OTHER strategies, characterized by the use of PastImpf indicative in just one of the linked clauses (cf. (8)-(10)).

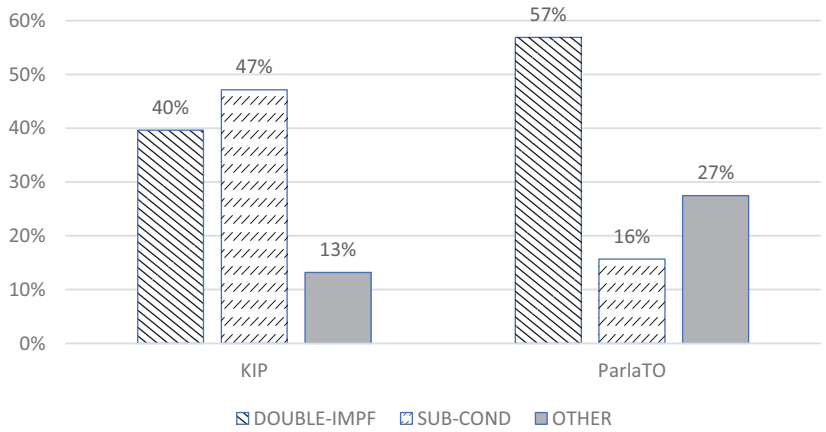

Figure 3: Counterfactual strategies in the KIP and ParlaTO modules.[6]

Since the use of PastImpf indicative characterizes both DOUBLE-IMPF and OTHER strategies, we can oppose the use of the SUB-COND strategy to all the strategies including at least one PastImpf indicative form (in the protasis, in the apodosis, in both – henceforth IND-IMPF), as shown in Figure 4:

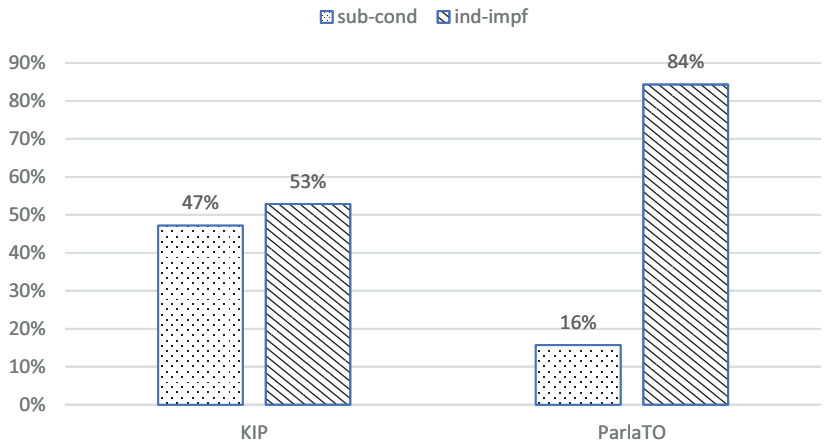

Figure 4: Distribution of SUB-COND and IND-IMPF strategies in KIP and ParlaTO.[7]

6 According to the Chi-Square Test, the result is significant at $p < .01$**, with $p =< 0.002094$, N=104.
7 According to the Fisher exact test, the result is significant at $p < .001$***, with $p =< 0.0007$, N=104.

As Figure 4 shows, the frequency of IND-IMPF counterfactual strategies, that is, strategies including one or two PastImpf indicative forms, is significantly higher in the ParlaTO corpus (84%), where the use of the SUB-COND strategy does not reach the 20% of the cases. Conversely, in the KIP corpus the distribution of SUB-COND strategies and IND-IMPF strategies is roughly even. How can we explain these data?

Given the socio-demographic differences between the two corpora (cf. §2.1), we can hypothesize that the most frequent use of the SUB-COND strategy in KIP is due to the higher ratio of educated speakers in that module and that, on the other hand, the greater frequency of the IND-IMPF strategy in the ParlaTO is due to the presence of a grater variation on the diastratic dimension, which would favor traits of popular and substandard Italian (Berruto 2012: 129; cf. Berruto 2014; Cerruti 2018). However, the difference between the two modules lies also on the diaphasic axis, so we cannot exclude that the differences are due to the presence of formal interactions in the KIP, which are instead completely absent in the ParlaTO, that is, to a difference of register. In the next section, we will answer this question, analyzing the distribution of the different constructions according to the parameters adopted in this study, focusing on the formality of the interaction (diaphasic indicator) and the educational qualification (diastratic indicator).

2.3 Diaphasic and diastratic variation

To test the statistical significance of our data, we employed the Random Forest method. The Random Forest method is useful in studies where the dataset is small, but the number of predictors is large. It is made up of several inference trees, therefore it has better predictive accuracy than a single tree, allowing to rank the predictors (Tagliamonte and Baayen 2012; Levshina 2015).

Let us first consider the distribution of SUB-COND *vs.* IND-IMPF constructions. As Figure 5 shows, the use of SUB-COND strategy reveals a significant relation with both the level of formality and the degree of the speakers' educational achievement, i.e. the two dots located in the right-hand part of the plot.

If we employ the Fisher Exact Test to verify the respective significance of these two predictors, we observe that the factor having the highest predictive value ($p < .001^{***}$) is the degree of educational achievement. In other words, speakers with higher educational achievement (i.e. higher on the diastratic axis) are more likely to convey counterfactual conditionals by means of past subjunctive followed by past conditional, than speakers with lower educational achievement (Figure 6). The latter tend to use instead counterfactual strategies including an ImpfPast indicative form (i.e. IND-IMPF strategy).

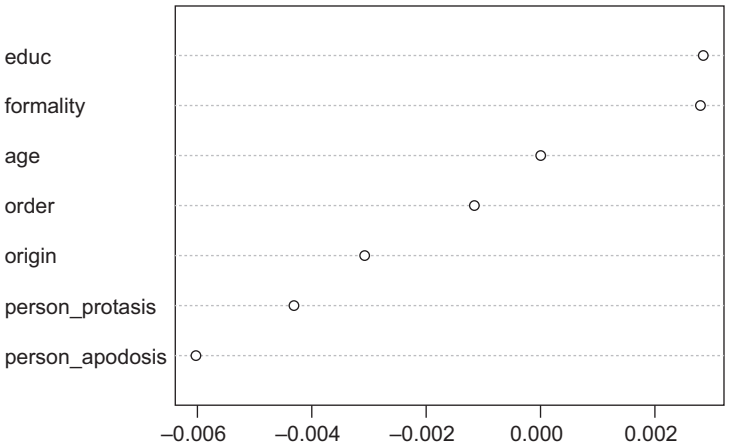

Figure 5: Plot ranking linguistic and extra-linguistic predictors for the occurrence of SUB-COND counterfactual strategies.[8]

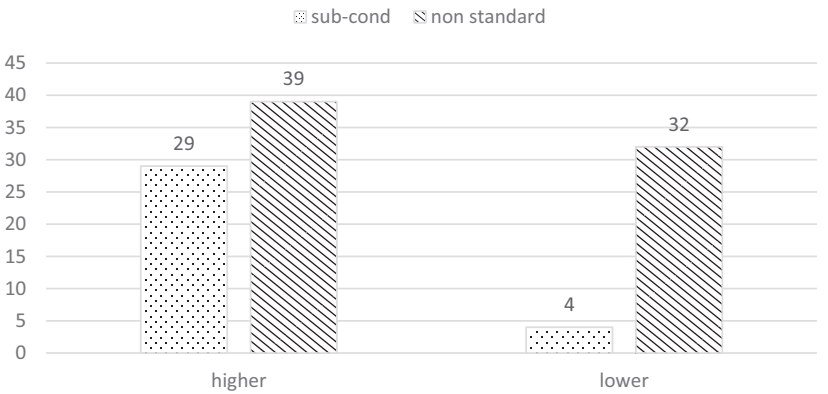

Figure 6: Use of SUB-COND and IND-IMPF strategies by higher educated and lower educated speakers (in this chart, the absolute numbers are provided).[9]

The formality of the interaction has a lower predictive value (p<.01**), but still significantly correlates with the type of counterfactual construction being used. As shown by Figure 7, in informal contexts, speakers are more likely to employ IND-IMPF strategies, thus recurring to counterfactual indicative forms, than in formal contexts, where the SUB-COND strategy prevails.

8 The model validity is excellent (Goodness of fit): C= 0.8014385.

9 According to the Fisher exact test, the result is significant at p < .001***, with p=< 0.0009, N=104.

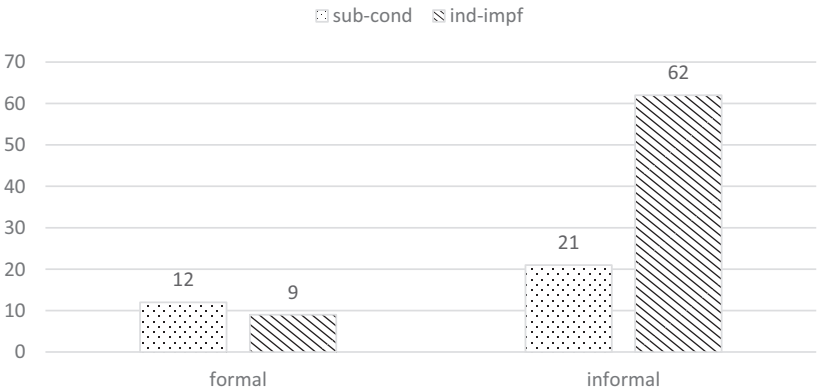

Figure 7: Use of SUB-COND and IND-IMPF strategies in formal and informal interactions speakers (in this chart, the absolute numbers are provided).[10]

Up to now, we opposed the SUB-COND strategy to the IND-IMPF strategy, which includes all the constructions in which a counterfactual PastImpf indicative occurs, either in isolation, followed or preceded by a subjunctive or a conditional, or in combination with another PastImpf indicative form. If we focus on the latter case, namely on the symmetric DOUBLE-IND strategy, we note that it does not show a significant correlation with formality, but only with the degree of educational achievement, which is the only predictor having statistical significance (Figure 8).

However, even for the degree of educational achievement, the Fisher Exact Test provides a low value of significance ($p < .05$*), correlating the use of the double PastImpf indicative to speakers with lower educational achievements, as shown in Figure 9. In the light of what we have seen in this section, we can confirm that the distribution of the different counterfactual strategies attested in Spoken Italian is mainly affected by the diastratic variation, in particular it can be linked to the level of educational achievement: a lower level of educational achievement strongly correlates with the use of counterfactual indicative forms (ImpfPast indicative), both in symmetric and asymmetric constructions, whereas a higher level of educational achievement correlates with the use of the asymmetric SUB-COND strategy, with past subjunctive in the protasis and past conditional in the apodosis. The use of counterfactual indicative also correlates, though with a weaker significance, to informal interactions, in which also educated speakers tend to recur to IND-IMPF strategies. We could not find such strong correlations explaining the distribution of symmetric constructions with double ImpfPast indicative (DOUBLE-IND), which appear to be

10 According to the Fisher exact test, the result is significant at $p < .01$**, with $p =< 0.0082$, N=104.

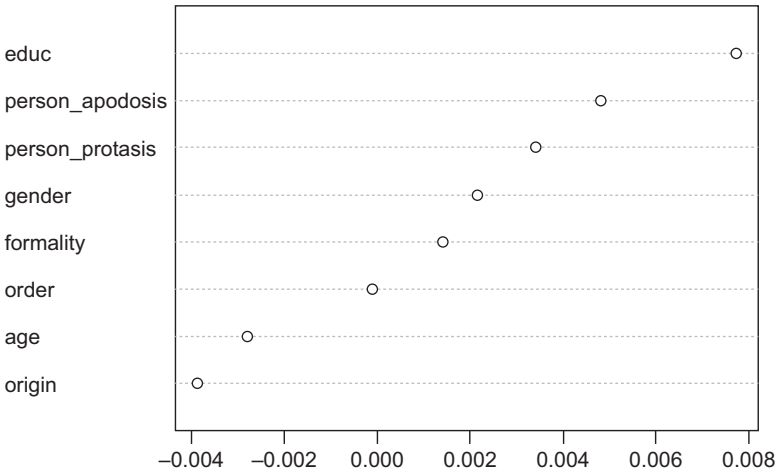

Figure 8: Plot ranking linguistic and extra-linguistic predictors for the occurrence of DOUBLE-IND counterfactual strategies.[11]

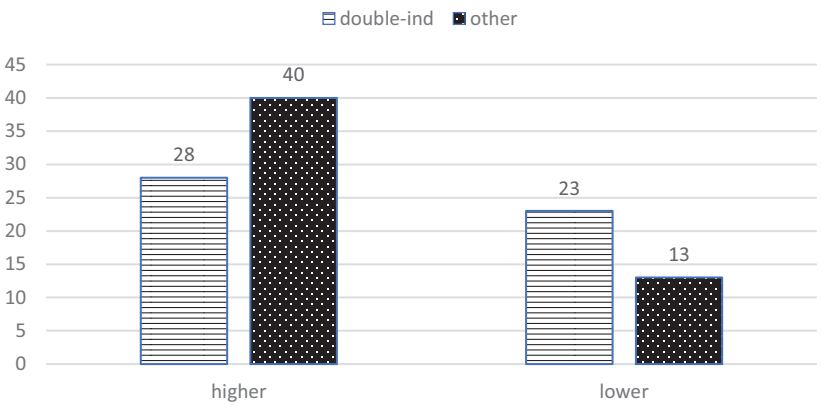

Figure 9: Use of DOUBLE-IND strategy by higher educated and lower educated speakers (in this chart, the absolute numbers are provided).[12]

only weakly connected with a lower degree of educational achievement. In other words, speakers appear to use the symmetric construction with double indicative widely in discourse, both in formal and informal contexts, though with a greater frequency in diastratically lower productions.

11 The model validity is excellent (Goodness of fit): C= 0.8047273.

12 According to the Fisher exact test, the result is significant at $p < .05$*, with $p =< 0.0389$, N=104.

To sum up, the observed correlations are revealing of a gradual expansion of counterfactual ImpfPast indicative in Italian, especially in symmetric constructions, starting from the lower varieties. How can we interpret these data in the light of a larger typological picture? Is the use of symmetric counterfactual constructions a recurrent tendency across languages, as argued by Haiman and Kuteva (2002), which may explain why the use of DOUBLE-IND in Italian appears to be widespread and not closely associated to a specific diastratic or diaphasic variety? And how frequent is the use of an indicative past imperfective form, typically employed for past habitual events and narrations, for counterfactual conditionals? In the next section, we will consider the attested cross-linguistic variation in the expression of counterfactual conditionals, looking for a larger picture, in the light of which the intra-linguistic variation described for Italian could be further explained.

3 Beyond Italian: Cross-linguistic variation in the expression of counterfactual conditionals

3.1 From sociolinguistics to typology: Methodological issues

As already argued in §1, we aim to integrate the analysis of the intra-linguistic variation attested in spoken Italian within a larger picture, to explain the rise and spread of nonstandard conditional constructions in the light of the attested cross-linguistic variation, verifying to what extent the patterns described for Italian in §2 follow or deviate from some general tendencies. In particular, we will focus on two aspects: the use of a symmetrical strategy for counterfactual conditionals and the use of a past imperfective form with counterfactual functions. To achieve this goal, we will compare the results obtained in the sociolinguistic analysis with both the variation attested in a variety 223-language sample and the existing typological literature on counterfactuals.

For each counterfactual construction, we focused on the verbal forms attested (or described) in counterfactual protasis and apodosis, their TAM properties, and on the asymmetry *vs*. symmetry of the construction. In addition, we monitored the presence of an independent dedicated counterfactual marker. For every language we further coded data regarding the strategies attested for hypothetical conditionals (cf. §1.1 for a definition).

We collected data from a variety sample composed of 223 languages (Mattiola 2020, freely available at http://amsacta.unibo.it/6504/). A variety sample is a language sample that aims at maximizing linguistic variety leaving aside (at least partially) the perfect statistical balancing. This kind of sample perfectly fits with

our final goal of verifying the existence of cross-linguistic patterns and tendencies. The criteria adopted by Mattiola (2020) to build this sample are the Diversity Value technique (Rijkhoff et al. 1993; Rijkhoff and Bakker 1998) and Ethnologue 2018 language classification (Lewis et al. 2018), as described in Table A.1 published as Appendix 1 in Miestamo et al. (2016) (for further information on this variety sample, see Mattiola 2020).

Our final sample is composed of 203 languages (see Appendix 1) because for some languages we could not have access to some grammatical descriptions. Among these 203 languages, we were able to identify information on counterfactual conditionals for 98 languages.

Before discussing the results of our analysis, it is necessary to highlight some crucial methodological issues that may have affected the study, which act as a big *caveat* in the interpretation of what we found. First of all, our sociolinguistic analysis focuses on spoken language data, without considering data from written language. A crucial issue with typological data from descriptive grammars is that in many cases we have no information about the specific variety that the grammar is describing: is it spoken language? Are we dealing with translations of written data? Field linguists indeed often recur to translations of sacred text (written), elicitation and translations (e.g., translations of sentences created ad hoc), monologues (songs, myths and traditional stories; or personal stories) which cannot account for the intra-linguistic variation of a language.

Furthermore, the sociolinguistic analysis of Italian is based on spontaneous speech, in both formal and informal contexts, but the data we usually find within descriptive grammars are hardly of this type. Most typological data discussed in descriptive grammars do likely not come from spontaneous interactions, but rather from more "controlled" situations, such as the ones involving the use of elicitation tecniques. This can be so for a number of possible reasons, e.g., the interviewer is not part of the informants' community, or there may be an asymmetrical relationship between them, just to name two examples. As a consequence, our typological findings are likely biased towards less spontaneous and informal data.

For an integrated approach as the one we are adopting in this study, the lack of information regarding the specific language variety being described, and, more in general, the sociolinguistically relevant properties of the recorded speakers and contexts of interaction, constitutes a crucial problem.

In comparing Italian data with the attested typological variation, we realized that in most cases we could not know what type of data we were dealing with. Yet, comparability is a crucial issue in typological research: it is indeed extremely important to compare data that can be thought as expressions of the (near-)same varieties, or at least varieties that can be compared among them. Unfortunately, though, data from descriptive grammars usually do not provide information regard-

ing the diastratic and/or diaphasic characterization of the data being described. We believe indeed that it would be important to re-discuss the overall methodology that stands behind fieldwork in order to make the data as informative as possible both for typologically-oriented and sociolinguistically-oriented comparative investigations, fostering the use of spontaneous interactions where possible.

3.2 Typological variation: Focus on symmetry and past tense

Data from the typological sample first of all confirm that the use of a symmetrical counterfactual construction is a widespread phenomenon, allowing us to analyze the DOUBLE-IND Italian strategy as a manifestation of a recurrent construction type. Out of the 98 languages in the sample that show some overtly TAM coding to mark counterfactual, 22 languages indeed show a symmetrical strategy, employing the same verb form for protasis and apodosis. For instance, in Mehek (Sepik), counterfactual conditionals are marked with the conditional morpheme *-na* in both protasis and apodosis,[13] while in Warupu (Skou), they are marked with the irrealis marker *n-*, as shown in the following examples.

(11) Mehek (Sepik; Hatfield 2016: 373)
loko *wate-**na*** *on* *ya-**na**-yun*
rain rain.fall-**COND** 1SG come-**COND**-1SG
'If it had rained, I would have come.'

(12) Warupu (Skou; Corris 2005: 340)
Kanro ***n**-e-n-ikoko,* *bârém* *beya* ***n**-o-te-ni* *vai.*
shoes **IRR-1SG.F-1SG.F**-wear thorn NEG **IRR-3SG.F**-shoot-1SG.F POL
'If I had worn shoes, the thorn would not have spiked me.'

The presence of symmetrical strategies is even more significant if we compare counterfactual to hypothetical conditionals, as shown in Figure 10. While asymmetrical constructions are the most common strategy attested in our sample, symmetrical constructions are more frequent for counterfactual conditionals than for possibility conditionals, to the point of representing almost a quarter of the languages.

13 Interestingly, if the result is a negative outcome, the verb in the apodosis doesn't have the conditional marker but only the negative makers *-nak* (Hatfield 2016: 372).

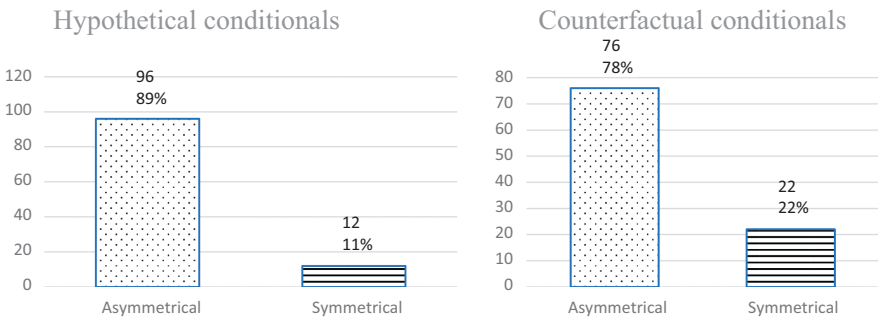

Figure 10: Symmetrical and asymmetrical constructions in counterfactual and hypothetical conditionals: data from the variety language sample.[14]

If we focus on the verbal forms employed in symmetrical constructions, interesting data emerge also from the range of markers used in both protasis and apodosis, as shown in Figure 11.

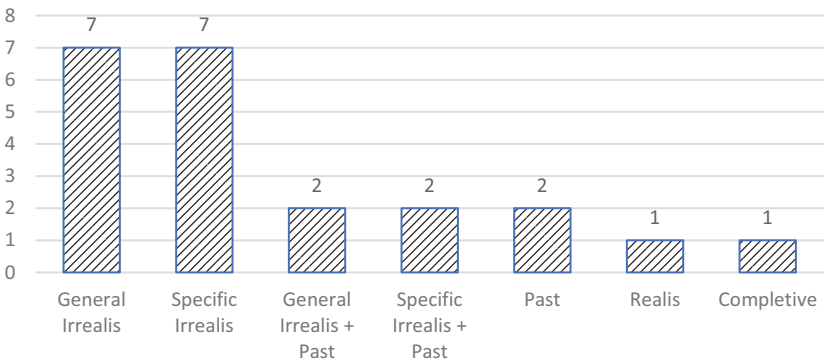

Figure 11: TAM markers in symmetrical counterfactual constructions.

Most symmetrical counterfactual constructions are marked by general irrealis markers, employed to denote a wide range of unrealized situations, or specific irrealis markers, typically mood markers employed for a restricted range of unrealized or potential situations:[15] in other words, in line with the results of Olguín Martínez

14 According to the Fisher exact test, the result is significant at p<.05*, with p=< 0.0381, N=205.

15 For a detailed discussion on the notion of irrealis and general irrealis marker, see Mauri and Sansò (2012). We classified as specific irrealis markers all those mood markers employed for a restricted range of unrealized or potential situations, including e.g., conditional markers, subjunctive mood markers, potential markers, hypothetical markers. There is great variation with respect to

and Lester's (2021) study, if the same verb form is used for protasis and apodosis, it is likely to be a form encoding some general or specific irrealis function (this is the case for 18 out of 22 languages showing symmetrical counterfactual strategies, cf. Figure 11). This could be due to an iconic mechanism, motivated by the double contrary-to-fact status of the two linked states of affairs. Whereas it is true that also for counterfactual conditionals the linked states of affairs stand in a consequential relation, which is proper of all conditional relations and motivates the use of asymmetric constructions, it is also true that in counterfactuals the consequential relation is less salient, in that it does not affect the truth values of the states of affairs, which are equally claimed to be contrary to fact.

In six languages, we observe the use of a symmetrical construction employing some past tense marker, either alone, as in example (13) from Tboli (Austronesian), or together with a specific or general irrealis marker, as in example (14) from Wardaman (Australian). In Tboli both protasis and apodosis are marked with *deng* which encodes a completed action:

(13) Tboli (Austronesian; Porter 1977: 153)
Ke ***deng*** *ton-em* *#* ***deng*** *benli-hem* *#* *ne.*
if **PAST** saw-you it **PAST** buy-you it now
'If you had seen it, you would have bought it.' (The fact is he hadn't seen it so he didn't buy it).

(14) Wardaman (Australian; Merlan 1994: 188)
bujun ***yi****-ngan-wo-****ndi*** *ma-jad* ***yi****-ngong-wo-****ndi***
if **IRR**-3SG/1SG-give-**PAST** MA-big.ABS **IRR**-2SG/1SG-give-**PAST**
ngawun, *ma-jad-wagbawun*
no MA-big-lacking.ABS
'If he had given me a lot, I would have given you [some], [but] no, [he did not] give a lot.'

The role of past tense in the coding of counterfactual conditionals appears to be far from marginal, especially if we consider also asymmetrical constructions: out of the 98 languages in the sample that show some overtly TAM coding to mark counterfactual, 28 languages use past tense to mark at least one between the protasis and the apodosis. For instance, in Jamsay Dogon (Dogon), counterfactuals are char-

the types of markers and the terminology through which they are referred to in descriptive grammars. This classification does not imply any theoretical claim, but is adopted here for descriptive purposes.

acterized by the use of the past particle *jì:ⁿ* in both protasis and apodosis,[16] while in Kayardild (Australian) the past suffix *-jarra* only marks the verb in the protasis, as shown in the following examples:

(15) Jamsay Dogon (Dogon; Heath 2008: 585)
nì-dî:ⁿ *yěs-sà-bà* ***jì:ⁿ*** *dèy,* *bé* *wǒ:-m̀* ***jì:ⁿ***
here come-RESLT-3PL.SBJ **PAST** if 3PL.OBJ kill.IMPF-1SG.SBJ **PAST**
'If they had come here, I'd have killed them.'

(16) Kayardild (Australian; Evans 1995: 260–261)
ngada *kurri-**jarra*** *bukaji-na* *dii-n-kina,* *ngada* *raa-ju*
1SG.NOM see-**PAST** seahawk-M.ABL sit-N-M.ABL 1SG.NOM spear-POT
'If I had seen a sea-hawk landing, I'd have speared it.'

Interestingly, in most cases, past tense is used in both protasis and apodosis (17 languages out of 28 languages employing past markers in counterfactual constructions). This does not necessarily imply that the construction is symmetrical: in some cases, even though the past marking is present in both verbs, protasis and apodosis can show other differences in the TAM marking. For instance, in Palauan (Austronesian), protasis and apodosis are both marked with the past tense, but the protasis also takes the hypothetical verb form (Josephs 1975: 447). This is the kind of situation that we also find in the SUB-COND strategy described in §2.2 for Italian, where the protasis is marked with a past subjunctive form and the apodosis with a past conditional form (i.e., forms that we would have classified here as specific irrealis).

The relevance of past tense for the coding of counterfactual conditionals is confirmed by the literature. This phenomenon has indeed been analyzed in different languages, such as in English by Langacker (1978), Proto-Uto-Aztecan (Steele 1975), Korean and Japanese (Han 1996; Cho 1997). Other studies have adopted a cross-linguistic perspective, noting that the presence of past tense markers for counterfactual situations is widespread in the world's languages (James 1982, Fleischman 1989). The correlation between past tense and counterfactuality was mostly explained by means of the so-called *past-as-unreal hypothesis* (cf. Dahl 1997). This hypothesis has different formulations, but the basic common premises are the following: past situations are perceived as distant from the realm of immediate realis (i.e., here-and-now) to the point of potentially being associated with unactualized situations (e.g., future events), and thus with irrealis. Therefore, it is argued that, due to some conceptual similarity, past tense markers can extend their functional

16 Occasionally *jì:ⁿ* can be omitted in the protasis (Heath 2008: 585).

domain to irrealis functions, like counterfactuality (cf. Steele 1975; Langacker 1978; James 1982; among others).

Dahl (1997: 100) points out a major problem with this theory, that is, the fact that the "marking of irreality/hypotheticality is rarely done by means of a past tense alone but normally by the combination of a past tense with something else". This "something else" is typically a marker with some modal meaning or a fully-fledged irrealis marker, as we saw for Wardaman in (14), which uses an irrealis prefix *yi-* and a past suffix *-ndi* to mark both verbs in protasis and apodosis.

In the light of the cross-linguistic variation attested in our language sample, we may thus analyze the two most frequent counterfactual strategies attested in Italian (SUB-COND and DOUBLE-IND) as manifestations of construction types that are recurrent across languages: the SUB-COND construction is an asymmetrical strategy composed of two different specific irrealis verb forms (subjunctive and conditional) inflected for the past tense; the DOUBLE-IND construction is a symmetrical strategy involving the use of a past indicative form, i.e., a verbal form that does *not* encode an irrealis function (cf. example (11) from Tboli). We may further add that also the OTHER constructions types, namely asymmetrical strategies involving the use of the past imperfective indicative form in just one of the linked clauses (cf. §2.2), follow a pattern that is attested in our languages sample, as exemplified by Kayardild (example (16)).

As argued in §2, the PastImpf indicative verbal form employed in nonstandard Italian strategies (that is, DOUBLE-IND and OTHER strategies), is characterized by an imperfective aspect which makes it the preferred strategy to refer to past habitual situations and narrations in the past. Yet, the data we collected in our typological sample do not provide clear patterns regarding aspectual features of the attested forms. Interestingly, as we saw in examples (4) and (5), PastImpf indicative is attested in Italian for a larger range of functions than the sole reference to past imperfective events, including unrealized situations such as fictitious or imaginary situations. May this function widening towards the irrealis domain play a role in the increasing use of PastImpf indicative in asymmetrical and symmetrical counterfactual strategies? The cross-linguistic patterns observed in our sample seem to suggest that the frequency of the DOUBLE-IND construction could be connected at least partially to the acquisition of a more general irrealis function by the PastImpf indicative form.

3.3 Past habituals as counterfactuals: A recurrent pattern

On the basis of a sample of a few dozen languages and focusing only on the marking of the counterfactual apodosis, Lazard (1998, 2006: 62) argues that the combination of some irrealis morpheme and a past tense morpheme is likely the more fre-

quent strategy in the languages of the world. However, he notes that the 'irrealis' morpheme "may assume different shades of meaning" (Lazard 2006: 62). So, for instance, in some languages, the irrealis morpheme is a future marker, like in Ditidaht (17), while in other languages we can find a desiderative, debitive, or potential marker, like in the case of Yimas (18).

(17) Ditidaht (Wakashan; James 1982: 382 citing Klokeid 1976: 31)
(ʔoyi) kab'at'p ***it*** *qwiy s,*
(if) know **PAST** specifier I
hitaqaya ***ʔi:k*** ***ibt*** *s ʔa:bay ʔoyi*
come **future** **PAST** I yesterday time-marker
'If I had known (it), then I would have come yesterday'

(18) Yimas (Ramu-Lower Sepik; Foley 1991: 265)
ampi*-ya-****ntuk****-mp-n* ***ant****-ka-wa-****ntut***
POT-come-**RM.PAST**-VII.SG-OBL **POT**-1SG.A-go-**RM.PAST**
'If those two had come, I would have gone.'

Another frequent situation noted by Lazard is the usage of past subjunctive or optative forms, which convey non-actualized events. This is, for instance, what we find in the counterfactual construction that is ascribed to the standard variety of Italian which, as noted in §2, uses a past subjunctive form in the protasis and a past conditional in the apodosis:

(19) Italian (Indo-European)
ovviamente ***se*** ***poi*** ***avesse*** ***partorito***
obviously if then AUX:SUBJ.PAST:3SG give.birth:PAST.PTCP
'obviously if then she had'
ci ***saremmo*** ***dovuti*** ***fermare*** *in ambulanza*
REFL.1PL AUX:COND:1PL must:PAST.PTCP:M.PL stop:INF in ambulance
'we should have stopped in an ambulance' [ParlaTO, PTA18]

Finally, Lazard also notes that, in some cases, counterfactuality can be marked by a combination of past marking with habitual marking. Habitual, however, does not refer to irrealis or unactualized situations, it rather conveys an aspectual value typically associated to actualized events: according to Comrie (1976: 30), habitual marking indeed denotes "any situation that can be protracted sufficiently in time, or that can be iterated a sufficient number of times over a long enough period". The correlation between habituals and counterfactual conditionals is addressed by Iatridou (2000, 2010), according to whom imperfective marking, or more pre-

cisely, habitual marking, is a crucial grammatical ingredient of counterfactual conditionals.

Italian is thus not isolate, in recurring to a past habitual form to express counterfactuality, but this strategy appears to be attested in several languages, independently of genealogical or areal proximity. For example, in Turkish, the apodosis of counterfactual conditionals is marked by combination of the aorist morpheme *-(ı)r* and the past tense morpheme *-dı*, as shown in the following example:

(20) Turkish (Turkic; Caro 2012: 20)
*Bir kedi-m ol-sa o-nu okşa-**r**, sev-**er**, o-ndan can al-**ır-dı**-m*
one cat-1SG be-COND it-ACC pet-**AOR** love-**AOR** it-ABL soul take-**AOR-PAST-1SG**
'If I had had a cat, I would have petted it, loved it, and taken its life.'

The Turkish aorist denotes continuing activity and can be used to mark habitual actions (but interestingly, aorist forms never express progressive aspect, Göksel and Kerslake 2005: 290; Lewis 1985: 117), for instance:

(21) Turkish (Turkic; Göksel and Kerslake 2005: 290)
*O zamanlarda Mehmet çok sigara iç-**er-di***
that at.times Mehmet very cigarette smoke-**AOR-PAST.COP**
'At that time Mehmet smoked/used to smoke a lot/Mehmet was a heavy smoker.'

The use of past habitual to mark counterfactual conditionals seems also quite frequent in Indo-Iranian languages (Lazard 1975, 1998). For instance, Pashto (Indo-European) has a "past narrative" formed by the particle *bə* and a finite verb form in the past, which usually expresses habitual events in the past but can also be used in the apodosis of counterfactual conditionals (Grjunberg 1987: 158–159; Vydrin 2011: 73), as shown in the following examples.

(22) Pashto (Indo-European; adapted from Vydrin 2011: 73)
ka obo wəṛay wāy no
if water take.PTCP.PAST.M be.OPT then
mor ta bə me dzawāb war-kāwə
mother POST PRTCL 1SG.ENCL.OBL answer PREF-do.PAST.3SG.M
'If the water had taken me, what would I have said to my mother.'

In Yaghnobi (Indo-European), counterfactuality is conveyed by the imperfect (also labelled as "past continuous"), which is formed by the verbal stem with special person/number affixes and the prefix *a-*, and is mainly used to convey past habitual (Xromov 1972: 32; Vydrin 2011: 72):

(23) Yaghnobi (Indo-European; Vydrin 2011: 72 citing Xromov 1972: 32)
a. *ax sahár ajáxišt*
he early.morning wake.up.IMPF.3SG
'He used to wake up early in the morning.'
b. *agár divári anǐžimišt aláks, yarm.avǐrimišt*
if outside go.out.IMPF.1SG walk.IMPF get.warm.IMPF.1SG
'If I had gone out for a walk, I would have gotten warm.'

Beyond Indo-Iranian languages, the correlation between counterfactual and habituality in the past is also attested in languages scattered in different areas of the world. Example (24) shows the case of Lavukaleve, an isolate language spoken in the Solomon Islands:

(24) Lavukaleve (Isolate; Terrill 2003: 436)
aka kini tataveua-re-a la o-e-sia-le
then ACT be.missing-NF-SG.F SG.F.ART 3SG.F.OBJ-SBD-do-POT
o-vea ma-me.
3SG.F.OBJ-know 3PL.SBJ-HAB
'If [anything] was missing, they would have known it (but it never was).'

Counterfactual conditionals in Lavukaleve are formed by marking the protasis with the potential marker *-le* while the apodosis appears to be marked by the habitual marker *-me*. The absence of a past tense marker is due to the fact that Lavukaleve has explicit ways to mark present and future tense but has no past tense marker (Terrill 2003: 324, 328).

Finally, another interesting case is the one found in Goemai (Afro-Asiatic). In Goemai counterfactual conditionals, the protasis is marked with the conditional marker *lá* while the apodosis is marked with the irrealis marker *t'óng* ('would'). Moreover, both protasis and apodosis are also marked with the morpheme *d'ín* which "probably originated in the close past tense particle *d'ín*" (Hellwig 2011: 463), as shown in (25).[17]

17 The conditional marker *lá* and the irrealis marker *t'òng* are used also to mark other types of conditionals (Hellwig 2011: 457).

(25) Goemai (Afro-Asiatic; Hellwig 2011: 463)
Óerém mén ***d'ín*** ***lá*** *là* *hèn=**d'ìn*** ***t'óng***
beans 1PL.POSS **PAST.CL** **COND** produce(SG) 1SG.S=**PAST.CL** **IRR**
póe yòe (. . .)
give 2SGF.O
'If our beans had produced (fruit) (but they didn't), I would have given (the fruit) (to) you.'

Interestingly, both *lá* and *t'óng* can also function as habitual markers, often used together in the same sentence, as shown in the following example:[18]

(26) Goemai (Afro-Asiatic; Hellwig 2011: 243)
[hèn]$_A$=dók ***lá*** *t'ém [póe gòe]$_o$* ***t'óng*** *gòepé (. . .)*
1SG.S=RM.PAST **HAB** tell give 2SG.M.O **HAB** THAT/WHEN
'in the past, I used to tell you that (. . .)'

To sum up, we can say that despite not being irrealis marker in a narrow sense, habitual markers seem to play an interesting role in the coding of counterfactuality, regardless of specific linguistic families or areas. More generally, the connection between habitual and the irrealis domain has been discussed by various scholars in typology (see for instance Lazard 1975; Chung and Timberlake 1985; Givón 1994; Elliott 2000), in the light of the fact that many languages of the world use the same construction to express both actualized habitual and unactualized events. This connection has been mainly explained postulating a conceptual connection between habituality and irrealis due to the fact that habitual does not relate to a specific action, but rather to a tendency to act. For instance, Givón (1994: 322) argues that habitual is a "hybrid modality" since it shares some features of realis (higher assertive certainty) and some of irrealis (lack of specific temporal reference, lack of specific evidence, non-referring NPs under its scope).

Based on a typological survey, Cristofaro (2004) further notes that "the association between habituality and irrealis takes place for past rather than present habituality" (Cristofaro 2004: 261). In other words, if habituality in the present is expressed through the same construction used also to express any type of unactualized situation, then it is also used to express habituality in the past, however the reverse does not hold (Cristofaro 2004: 262–263). To explain this tendency, Cristofaro argues that once again there is a conceptual correlation between irrealis and

18 Hellwig (2011: 324) comments on this functional extension by proposing the following diachronic development CONDITIONAL > HABITUAL.

habituality in the past (as opposed to habituality in the present). Indeed, as already noted, both habitual events and past events individually have been connected conceptually to unactualized events (irrealis). Therefore, not only habituality in the past relates to situations that do not pertain to any specific actualized instance of the relevant event (like habitual in general), but it also does not hold in the present anymore and is thus perceived as far from the realm of immediate realis (like past in general).

As already noted, in Italian, indicative imperfective past is employed to denote habitual states or events in the past, but it has also developed sort-of irrealis functions related to the domain of imagination or possibility, where the subjunctive or conditional would be required in the standard variety, thus confirming Cristofaro's (2004) account. From this point of view, the widespread use of PastImpf indicative in Italian counterfactual constructions (as in (4), repeated here for convenience as (27)) is to be interpreted as the spread of a strategy employed for both past habituality and unactualized events.

(27) Italian (Indo-European)

Ah	*la*	*rossi*	*ci*	***faceva il culo***	*soltanto*	*se*	*lo*
ah	DEF	Rossi	1PL.DAT	scold:IND.IMPF.PAST:3SG	only	if	it

pensavamo
think:IND.IMPF.PAST:3SG
'ah, the Rossi would have scolded us if we had even thought' [KIP, TOA3001]

In the light of the typological variation discussed in this section, Italian is thus not to be seen as an exception to Olguín Martínez and Lester's (2021) predictions: the observed preference for the double use of PastImpf indicative, leading to a symmetric counterfactual strategy, indeed confirms the cross-linguistic tendency towards counterfactual constructions in which both protasis and apodosis are marked as unactualized.

4 Concluding remarks

In this contribution, our main aim was to provide a corpus-based and sociolinguistically-oriented description of counterfactual conditional constructions in spoken Italian, comparing our findings to the more general cross-linguistic variation.

First, through a corpus-based analysis, we identified DOUBLE-IMPF and SUB-COND as the most frequent strategies in Italian to convey counterfactual conditionals, as expected, but alongside these, we were also able to identify some other strategies

that are not well depicted in the literature, such as PastImpf Indicative + Past Conditional or Past Subjunctive + PastImpf Indicative. The distribution of the two main strategies is furthermore not balanced in the two corpora examined: while in the KIP corpus, SUB-COND was the most frequent strategy (even though with close percentages), DOUBLE-IMPF was by the far the most frequent in the ParlaTO corpus. The picture is even more unbalanced if we compare the distribution of SUB-COND to the distribution of IND-IMPF strategies, that is, strategies employing PastImpf indicative either in the protasis or in protasis, or in both. On the one hand, KIP shows almost the same percentage of SUB-COND and IND-IMPF strategies (47% *vs*. 53%), on the other hand ParlaTO shows a clear preference for IND-IMPF (84% *vs*. 16%). The reasons behind this imbalance can be traced back to sociolinguistic variables relating to the diaphasic and diastratic axes. SUB-COND strategies are clearly preferred by speakers with higher education and also in formal interactions, while conversely IND-IMPF strategies (especially the symmetric ones) are more often found in informal situations and used by less educated speakers, but also not fully dis-preferred in high communicative situations and by highly educated speakers, too. We can thus presume that counterfactual ImpfPast indicative is expanding its domain of application from nonstandard varieties of Italian.

Second, we tested our language-specific findings from a larger cross-linguistic perspective. From a methodological point of view, integrating data from sociolinguistic and from cross-linguistic investigations is far from being an easy task. The two kinds of data are indeed quite different from each other and can hardly be compared, at least until grammars will thoroughly provide sociolinguistic information about the variety described within. However, the information grammars offer can be useful as well, even though to a lesser extent, as a general litmus test for language-specific linguistic change. Our cross-linguistic investigation of counterfactual conditional constructions confirmed the tendency of Italian strategies for a preference for displaying symmetrical constructions and the frequent presence of past verbal forms.

Finally, we provided some typological considerations on why past imperfective forms are attested to express irrealis functions, such as counterfactual conditional. The fact that past imperfective forms are very often used to convey past habituality indeed leads them to lose their reference to individual and specific situations. Due to this weakening of their referential force, past habituals become conceptually similar to unactualized events, frequently extending their functional domain to irrealis functions. Under this respect, we propose to analyze the spread of a past imperfective form to counterfactual functions in Italian as a manifestation of this more general tendency.

Abbreviations

1,2,3	1st, 2nd, 3rd person
A	subject of transitive verb
ABL	ablative
ABS	absolutive
ACC	accusative
ACT	action (particle)
AOR	aorist
ART	definite article
AUX	auxiliary
CL	close (Past)
COND	conditional
COP	copula
DAT	dative
DEF	definite
ENCL	enclitic
F	feminine
HAB	habitual
IMP	imperative
IMPF	imperfective
IND	indicative
INDEF	indefinite
INF	infinitive
IRR	irrealis
LOC	locative
M	masculine
MA	adjective class
N	nominalization
NEG	negative
NF	non-Finite (verb suffix)
NOM	nominative
O	object pronoun
OBJ	object
OBL	oblique
OPT	optative
PAST	past
PL	plural
POL	polite
POSS	possessive
POST	postposition
POT	potential
PREF	prefix
PRES	present
PRTCL	particle
PTCP	participle

REFL	reflexive
RM	remote (Past)
RESLT	resultative
S	subject (intransitive and transitive) pronoun
SBJ	subject
SBD	subordinate (verb prefix)
SG	singular
SUBJ	subjunctive
VII	class VII

Appendix 1: 203-language sample (based on Mattiola 2020)

Language	Macro-area	Family	Sub-classification
Tamasheq	Africa	Afro-Asiatic	Berber
Buwal	Africa	Afro-Asiatic	Chadic
Goemai	Africa	Afro-Asiatic	Chadic
Hamer-Banna	Africa	Afro-Asiatic	Omotic
Kabuverdianu	Africa	Creole	Portuguese
Khwedam	Africa	Khoe-Kwadi (Khoisan)	Khoe
Bullom So	Africa	Niger-Congo	Atlantic-Congo
Noone	Africa	Niger-Congo	Atlantic-Congo
Chuwabu	Africa	Niger-Congo	Atlantic-Congo
Eton	Africa	Niger-Congo	Atlantic-Congo
Ngemba	Africa	Niger-Congo	Atlantic-Congo
Mungbam	Africa	Niger-Congo	Atlantic-Congo
Obolo	Africa	Niger-Congo	Atlantic-Congo
Yoruba	Africa	Niger-Congo	Atlantic-Congo
Mbembe Tigon	Africa	Niger-Congo	Atlantic-Congo
Cicipu	Africa	Niger-Congo	Atlantic-Congo
Gbari	Africa	Niger-Congo	Atlantic-Congo
Lijili	Africa	Niger-Congo	Atlantic-Congo
Jamsay Dogon	Africa	Niger-Congo	Atlantic-Congo
Godié	Africa	Niger-Congo	Atlantic-Congo
Ikposo	Africa	Niger-Congo	Atlantic-Congo
Mundang	Africa	Niger-Congo	Atlantic-Congo
Bouna Kulango	Africa	Niger-Congo	Atlantic-Congo
Lumun	Africa	Niger-Congo	Kordofanian
Maan	Africa	Niger-Congo	Mande

(continued)

Language	Macro-area	Family	Sub-classification
Ik	Africa	Nilo-Saharan	Kuliak
Zaghawa	Africa	Nilo-Saharan	Saharan
Gumuz	Africa	Nilo-Saharan	Satellite-Core
Lango	Africa	Nilo-Saharan	Satellite-Core
Ma'di	Africa	Nilo-Saharan	Satellite-Core
Koyra Chiini Songhay	Africa	Nilo-Saharan	Songhay
N|u	Africa	Tuu	!Ui
Gooniyandi	Australia & New Guinea	Australian	Bunaban
Wardaman	Australia & New Guinea	Australian	Gunwingguan
Bardi	Australia & New Guinea	Australian	Nyulnyulan
Warungu	Australia & New Guinea	Australian	Pama-Nyungan
Kayardild	Australia & New Guinea	Australian	Tangic
Worrorra	Australia & New Guinea	Australian	Worrorran
Garrwa	Australia & New Guinea	Australian	Yanyi
Imonda	Australia & New Guinea	Border	Waris
Lavukaleve	Australia & New Guinea	Central Solomons	//
Moskona	Australia & New Guinea	East Bird's Head-Sentani	East Bird's Head
Mali	Australia & New Guinea	East New Britain	Baining
Kuot	Australia & New Guinea	Isolate	//
Kwomtari	Australia & New Guinea	Kwomtari	Nuclear Kwomtari
Warembori	Australia & New Guinea	Lower Mamberamo	//
Yaul	Australia & New Guinea	Mongol-Langam	//
Nimboran	Australia & New Guinea	Nimboran	//
Rotokas	Australia & New Guinea	North Bougainville	Rotokas
Motu, Hiri	Australia & New Guinea	Pidgin	Motu
Yimas	Australia & New Guinea	Ramu-Lower Sepik	Lower Sepik
Mehek	Australia & New Guinea	Sepik	Tama
Warupu	Australia & New Guinea	Skou	Skou-Serra-Piore
Siwai	Australia & New Guinea	South Bouginville	Buin
Wára	Australia & New Guinea	South-Central Papuan	Morehead-Upper Maro
Yapunda	Australia & New Guinea	Torricelli	Wapei-Palei
Asmat, Central	Australia & New Guinea	Trans-New Guinea	Amsat-Kamoro
Menya	Australia & New Guinea	Trans-New Guinea	Angan
Duna	Australia & New Guinea	Trans-New Guinea	Duna-Bogaya
Kwewa, East	Australia & New Guinea	Trans-New Guinea	Engan
Yagaria	Australia & New Guinea	Trans-New Guinea	Kainantu-Goroka
Mauwake	Australia & New Guinea	Trans-New Guinea	Madang
Marind	Australia & New Guinea	Trans-New Guinea	Marind

(continued)

Language	Macro-area	Family	Sub-classification
Mian	Australia & New Guinea	Trans-New Guinea	Ok-Awyu
Oksapmin	Australia & New Guinea	Trans-New Guinea	Oksapmin
Daga	Australia & New Guinea	Trans-New Guinea	Southeast Papuan
Makasae	Australia & New Guinea	Trans-New Guinea	West
Tidore	Australia & New Guinea	West Papuan	North Halmahera
Aka-Jeru	Eurasia	Andamanese	Great Andamanese
Chukchi	Eurasia	Chukotko-Kamchatkan	Northern
Armenian	Eurasia	Indo-European	Armenian
Lithuanian	Eurasia	Indo-European	Balto-Slavic
Breton	Eurasia	Indo-European	Celtic
Frisian, Northern	Eurasia	Indo-European	Germanic
Greek	Eurasia	Indo-European	Greek
Palula	Eurasia	Indo-European	Indo-Iranian
Catalan	Eurasia	Indo-European	Italic
Basque	Eurasia	Isolate	//
Burushaski	Eurasia	Isolate	//
Laz	Eurasia	Kartvelian	Zan
Korean	Eurasia	Koreanic	//
Bonan	Eurasia	Mongolic	Eastern
Hinukh	Eurasia	North Caucasian	East Caucasian
Udihe	Eurasia	Tungusic	Southern
Uyghur	Eurasia	Turkic	Eastern
Nenets	Eurasia	Uralic	Samoyed
Ket	Eurasia	Yeniseian	//
Southern Yukaghir	Eurasia	Yukaghir	//
Arapaho	North America	Algic	Algonquian
Pawnee	North America	Caddoan	Caddoan
Ineseño	North America	Chumash	Chumashan
Havasupai-Walapai-Yavapai	North America	Cochimí-Yuman	Yuman
Yupik, Central	North America	Eskimo-Aleut	Eskimo
Chilcotin	North America	Eyak-Athabaskan	Athabaskan
Haida, Northern/ Southern	North America	Haida	//
Huave, San Dioniso del Mar	North America	Huavean	//
Karok	North America	Isolate	//
Klamath-Modoc	North America	Isolate	//

(continued)

Language	Macro-area	Family	Sub-classification
Kutenai	North America	Isolate	//
Molale	North America	Isolate	//
Tonkawa	North America	Isolate	//
Washo	North America	Isolate	//
Yuchi	North America	Isolate	//
Zuni	North America	Isolate	//
Kiowa	North America	Kiowa-Tanoan	//
Maidu, Northwest	North America	Maiduan	//
Itza'	North America	Mayan	Yucatecan-Core Mayan
Ohlone, Southern	North America	Miwok-Costanoan	Costanoan
Popoluca, Highland	North America	Mixe-Zoquean	Zoquean
Muskogee	North America	Muskogean	Muskogean
Zapotec, Coatlán / San Baltazar Loxicha	North America	Otomanguean	Eastern Otomanguean
Pomo, Southern	North America	Pomoan	Western
Yakama	North America	Sahaptian	Sahaptin
Halkomelem	North America	Salish	Central Salish
Hidatsan	North America	Siouan-Catawban	Siouan
Purepecha, Western Highland	North America	Tarascan	//
Chontal, Lowland Oaxaca	North America	Tequistlatecan	//
Tepehua, Huehuetla	North America	Totonacan	Tepehua
Nisga'a	North America	Tsimshian	Nass-Gitksan
Ute-Southern Paiute	North America	Uto-Aztecan	Northern Uto-Aztecan
Nuu-chah-nulth	North America	Wakashan	Southern Wakashan
Wintu	North America	Wintuan	//
Witoto, Murui	North America	Witotoan	Proto-Huitoto-Ocaina
Yuki	North America	Yukian	Core Yukian
Jamamadí	South America	Arauan	Jamamadi
Awa-Cuaquier	South America	Barbacoan	Northern
Trió	South America	Cariban	Tiriyó
Pakaásnovos	South America	Chapacuran	Wari
Kuna, Border	South America	Chibchan	Chibchan B
Uru	South America	Chipaya-Uru	//

(continued)

Language	Macro-area	Family	Sub-classification
Emberá, Northern	South America	Chocoan	Emberá
Cholón	South America	Cholonan	//
Tehuelche	South America	Chon	//
Guahibo	South America	Guajiboan	Guajibo
Pilagá	South America	Guaykuruan	Southern
Kwaza	South America	Isolate	//
Movima	South America	Isolate	//
Trumai	South America	Isolate	//
Urarina	South America	Isolate	//
Apinayé	South America	Jean	Northern
Wampís	South America	Jivaroan	Jívaro
Karajá	South America	Karajá	//
Katukína (/ Kanamarí)	South America	Katukinan	//
Qawasqar	South America	Kaweskaran	//
Tariana	South America	Maipurean	Northern
Mapudungun	South America	Mapudungu	//
Sanapaná	South America	Mascoyan	//
Maxakalí	South America	Maxakalian	//
Mískito	North America	Misumalpan	//
Tsimané	South America	Mosetenan	//
Mamaindé	South America	Nambikwara	Nambikwara Complex
Páez	South America	Paezan	Paezan
Kakataibo-Kashibo	South America	Panoan	Mainline
Hupdë	South America	Puinavean	Hupda
Quechua, Yauyos	South America	Quechuan	Central Quechua
Maco	South America	Sálivan	Piaroa-Maco
Cavineña	South America	Tacanan	//
Tanimuca-Retuarã	South America	Tucanoan	Western Tucanoan
Ninam	South America	Yanomaman	//
Semelai	Southeast Asia & Oceania	Austro-Asiatic	Mon-Khmer
Stieng, Bulo	Southeast Asia & Oceania	Austro-Asiatic	Mon-Khmer
Pnar	Southeast Asia & Oceania	Austro-Asiatic	Mon-Khmer
Kharia	Southeast Asia & Oceania	Austro-Asiatic	Munda
Atayal	Southeast Asia & Oceania	Austronesian	Atayalic
Bunun	Southeast Asia & Oceania	Austronesian	Bunun
Siraya	Southeast Asia & Oceania	Austronesian	East Formosan
Bali	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian

(continued)

Language	Macro-area	Family	Sub-classification
Ibatan	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Tboli	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Tukang Besi North	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
To'abaita	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Pampangan	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Chamorro	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Bajau, West Coast	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Cebuano	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Javanese	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Bakati'	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Madura	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Malay, Jambi	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Tondano	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Moken	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Ida'an	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Iraya	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Ilocano	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Nias	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Palauan	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Makasar	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Subanen, Northern	Southeast Asia & Oceania	Austronesian	Malayo-Polynesian
Paiwan	Southeast Asia & Oceania	Austronesian	Paiwan
Payuma	Southeast Asia & Oceania	Austronesian	Payuma
Rukai	Southeast Asia & Oceania	Austronesian	Rukai
Saaroa	Southeast Asia & Oceania	Austronesian	Tsouic
Thao	Southeast Asia & Oceania	Austronesian	Western Plains
Miao, Western Xiangxi	Southeast Asia & Oceania	Hmong-Mien	Hmongic
Mai Brat	Southeast Asia & Oceania	Maybrat	//
Dera	Southeast Asia & Oceania	Senagi	//
Mandarin Chinese	Southeast Asia & Oceania	Sino-Tibetan	Chinese
Naga, Ao	Southeast Asia & Oceania	Sino-Tibetan	Tibeto-Burman
Karbi	Southeast Asia & Oceania	Sino-Tibetan	Tibeto-Burman
Katso	Southeast Asia & Oceania	Sino-Tibetan	Tibeto-Burman
Gahri	Southeast Asia & Oceania	Sino-Tibetan	Tibeto-Burman
Yakkha	Southeast Asia & Oceania	Sino-Tibetan	Tibeto-Burman
Lao	Southeast Asia & Oceania	Tai-Kadai	Kam-Tai

References

Berretta, Monica. 1993. Morfologia. In Alberto A. Sobrero (ed.), *Introduzione all'italiano contemporaneo. Vol. I*, 193–245. Roma/Bari: Laterza.

Berruto, Gaetano. 1983. L'italiano popolare e la semplificazione linguistica. *Vox Romanica* 42. 38–79.

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

Berruto, Gaetano. 2014. Esiste ancora l'italiano popolare? Una rivisitazione. In Paul Danler & Christine Konecny (eds.), D*all'architettura della lingua italiana all'architettura linguistica dell'Italia. Saggi in omaggio a Heidi Siller-Runggaldier*, 277–290. Frankfurt am Main: Peter Lang.

Bertinetto, Pier Marco. 1986. *Tempo, aspetto e azione nel verbo italiano. Il sistema dell'indicativo*. Firenze: Accademia della Crusca.

Caro, Ryan. 2012. Semantics of the Turkish non-past. Unpublished manuscript, Yale University. Retrieved from: https://ling.yale.edu/sites/default/files/files/alumni%20senior%20essays/CaroSeniorEssay.pdf

Cerruti, Massimo. 2018. Il parlato regionale oggi: un italiano composito? *LId'O Lingua italiana d'oggi* 15. 15–31.

Cho, Eun. 1997. Counterfactuals in Korean and Japanese: Interaction between verbal morphology and interpretation. Unpublished manuscript, Cornell University, Ithaca, NY.

Chung, Sandra & Alan Timberlake. 1985. Tense, aspect, and mood. In Timothy Shopen (ed.), *Language Typology and Syntactic Description, Volume 3: Grammatical categories and the lexicon*, 202–258. Cambridge: Cambridge University Press.

Comrie, Bernard. 1976. *Aspect*. Cambridge: Cambridge University Press.

Comrie, Bernard. 1986. Conditionals: A typology. In Elisabeth Closs Traugott (ed.), *On Conditionals*, 77–99. Cambridge: Cambridge Univerity Press.

Corris, Miriam. 2005. *A grammar of Barupu, a language of Papua New Guinea*. Sydney: University of Sydney dissertation.

Crewe, William J. 1984. *Singapore English and Standard English: Exercises in awareness*. Singapore: Eastern Universities Press.

Cristofaro, Sonia. 2004. Past habituals and irrealis. In Yuri A. Lander, Vladimir A. Plungian &. Anna Yu Urmanchieva (eds.), *Irrealis and Irreality*, 256–272. Moscow: Gnosis.

Dahl, Östen. 1997. The relation between past time reference and counterfactuality: A new look. In Angeliki Athanasiadou & René Dirven (eds.), *On Conditionals Again*, 97–114. Amsterdam/Philadelphia: John Benjamins.

Elliott, Jennifer R. 2000. Realis and irrealis: Forms and concepts of the grammaticalization of reality. *Linguistic Typology* 4(1). 55–90.

Evans, Nicholas. 1995. *A grammar of Kayardild*. Berlin/New York: Mouton de Gruyter.

Fleischman, Suzanne. 1989. Temporal distance: A basic linguistic metaphor. *Studies in Language* 13(1). 1–50.

Foley, William. 1991. *The Yimas language of Papua New Guinea*. Stanford, CA: Stanford University Press.

Givón, Talmy. 1994. Irrealis and the subjunctive. *Studies in Language* 18(2). 265–337.

Göksel, Aslı & Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. London: Routledge.

Grjunberg, Aleksandr L. 1987. *Ocherk grammatiki afganskogo jazyka (pashto) [Grammar of the Afghan language (Pashto)]*. Leningrad: Nauka.

Haiman, John & Tania Kuteva. 2002. The symmetry of counterfactuals. In Joan L. Bybee & Michael Noonan (eds.), *Complex sentences in grammar and discourse: Essays in honor of Sandra A. Thompson*, 101–124. Amsterdam/Philadelphia: John Benjamins.

Han, Chung-hye. 1996. Comparing English and Korean counterfactuals: The role of verbal morphology and lexical aspect in counterfactual interpretation. In Anthony D. Green & Virginia Montapanyane (eds.), *Proceedings of the Eastern States Conference on Linguistics '96*, 124–138. Ithaca, NY: Cornell University.

Hatfield, Adam. 2016. *A grammar of Mehek*. Buffalo, NY: State University of New York at Buffalo dissertation.

Heath, Jeffrey. 2008. *A grammar of Jamsay*. Berlin/New York: Mouton de Gruyter.

Hellwig, Birgit. 2011. *A Grammar of Goemai*. Berlin/New York: Mouton de Gruyter.

Iatridou, Sabine. 2000. The grammatical ingredients of counterfactuality. *Linguistic Inquiry* 31(2). 231–270.

Iatridou, Sabine. 2010. Some thoughts about the imperfect in counterfactuals. Unpublished manuscript, MIT.

James, Deborah. 1982. Past tense and the hypothetical: A cross-linguistic study. *Studies in Language* 6(3). 375–403.

Josephs, Lewis S. 1975. *Palauan reference grammar*. Honolulu, HI: University of Hawaii Press.

Klokeid, Terry J. 1976. *Topics in Lardil grammar*. Cambridge, MA: MIT dissertation.

Langacker, Ronald W. 1978. The form and meaning of the English auxiliary. *Language* 54(4). 853–882

Lazard, Gilbert. 1975. La catégorie de l'éventuel. In René Amacker (ed.), *Mélanges linguistiques offerts à Émile Benveniste*, 347–358. Paris: Peeters.

Lazard, Gilbert. 1998. L'expression de l'irréel: essai de typologie. In Leonid Kulikov & Heinz Vater (eds.), *Typology of verbal categories: Papers presented to Vladimir Nedjalkov on the occasion of his 70th birthday*, 237–247. Tübingen: Niemeyer.

Lazard, Gilbert. 2006. More on counterfactuality, and on categories in general. *Linguistic Typology* 10(1). 61–66.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins.

Lewis, Geoffrey L. 1985. *Turkish grammar*. New York, NY: Oxford University Press.

Lewis, M. Paul, Gary F. Simons & Charles D. Fennig (eds.). 2018. *Ethnologue: Languages of the world. 21st edn*. Dallas, TX: SIL International. http://www.ethnologue.com/

Mattiola, Simone. 2020. *Two language samples for maximizing linguistic variety*. Bologna: AMS Acta.

Mauri, Caterina & Andrea Sansò. 2012. What do languages encode when they encode reality status? *Language Sciences* 34(2). 99–106.

Mauri, Caterina & Andrea Sansò. 2016. The linguistic marking of (ir)realis and subjunctive. In Jan Nuyts & Johan van der Auwera (eds.), *The Oxford Handbook of mood and modality*, 166–195. Oxford: Oxford University Press.

Mauri, Caterina, Silvia Ballarè, Eugenio Goria, 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*. https://ceur-ws.org/Vol-2481/paper45.pdf

Mazzoleni, Marco. 2013. Le concordanze dei periodi ipotetici tra italiano antico, dialetti italoromanzi ed italiano contemporaneo, *inTRAlinea* [=Special Issue *Palabras con aroma a mujer. Scritti in onore di Alessandra Melloni.* Edited by Maria Isabel Fernández García & Mariachiara Russo].

Merlan, Francesca C. 1994. *A grammar of Wardaman: A language of the Northern Territory of Australia.* Berlin/New York: Mouton de Gruyter.

Miestamo, Matti, Dik Bakker & Antti Arppe. 2016. Sampling for variety. *Linguistic Typology* 20(2). 233–296.
Olguín Martínez, Jesús & Nicholas Lester 2021. A quantitative analysis of counterfactual conditionals in cross-linguistic perspective. *Italian Journal of Linguistics* 33(2). 147–182.
Porter, Doris. 1977. *A Tboli grammar*. Manila: Linguistic Society of the Philippines.
Rijkhoff, Jan, Dik Bakker, Kees Hengeveld & Peter Kahrel. 1993. A method of language sampling. *Studies in Language* 17(1). 169–203.
Rijkhoff, Jan & Dik Bakker. 1998. Language sampling. *Linguistic Typology* 2(3). 263–314.
Schachter, Jacqueline C. 1971. *Presupposition and counterfactual conditional sentences.* Los Angeles, CA: University of California Los Angeles dissertation.
Steele, Susan. 1975. Past and irrealis: Just what does it all mean? *International Journal of American Linguistics* 41(3). 200–217.
Sweetser, Eve E. 1990. *From etymology to pragmatics*. Cambridge: Cambridge University Press.
Tagliamonte, Sali A. & Harald Baayen. 2012. Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178.
Taylor, John R. 1997. Conditionals and polarity. In Angeliki Athanasiadou & René Dirven (eds.), *On conditionals again*, 289–306, Amsterdam/Philadelphia: John Benjamins.
Terrill, Angela. 2003. *A grammar of Lavukaleve*. Berlin/New York: Mouton de Gruyter.
Thompson, Sandra A., Robert E. Longacre & Shin Ja J. Hwang. 2007. Adverbial clauses. In Timothy Shopen (ed.), *Language typology and syntactic description. Vol.2: Complex constructions*, 237–300. 2nd edn. Cambridge: Cambridge University Press.
Vydrin, Arseniy. 2011. Counterfactual mood in Iranian. In Agnes Korn, Geoffrey Haig, Simin Karimi & Pollet Samvelian (eds.), *Topics in iranian linguistics*, 71–88. Wiesbaden: Reichert.
Wiberg, Eva. 2010. Imperfetto. In Raffaele Simone, Gaetano Berruto & Paolo D'Achille (eds.), *Enciclopedia dell'Italiano*. Roma: Istituto della Enciclopedia Italiana Treccani. http://www.treccani.it/enciclopedia/imperfetto_(Enciclopedia-dell%27Italiano)/
Xromov, Albert L. 1972: *Jagnobskij jazyk.* Moscow: Nauka.
Zaefferer, Dietmar. 1991. Conditionals and unconditionals: Cross-linguistic and logical aspects. In Dietmar Zaefferer (ed.), *Semantic Universals and Universal Semantics*, 210–236. Berlin/New York: Mouton de Gruyter.
Ziegeler, Debra. 1994. Conditionals and counterfactuality in Singaporean English. *Journal of Intercultural Studies* 15(1). 29–49.