

Article

Italian Discourse Markers in First- and Second-Generation Migrants in Naples: The Case of *Cioè*

Violetta Cataldo 

Department of Humanities, University of Naples Federico II, 80133 Naples, Italy; vcataldo@unisa.it

Abstract

This study investigates the use of the discourse marker *cioè* in the Italian speech of first- and second-generation migrants living in Naples, focusing on frequency, phonetic realization, pragmatic functions, and syntactic position. While previous research on discourse markers in Italian as a second language has primarily addressed their functional development, their phonetic realization—especially in contact with regional varieties—remains largely unexplored. The analysis is based on a corpus of sociolinguistic interviews collected within the HELLO Campania! Project and includes Sri Lankan and Filipino speakers, compared with a control group of native Neapolitan speakers. Quantitative analyses reveal clear generational differences: first-generation speakers show low frequencies and limited functional range, whereas second-generation speakers display native-like frequency and distribution. Crucially, results show that migrants not only acquire *cioè* at a functional level but also reproduce locally salient phonetic variants, including variety-specific patterns of phonetic reduction, i.e., [tʃɛ] and [ʃɛ]. These findings suggest that linguistic integration operates across multiple levels and that discourse markers are sensitive sites of socio-phonetic accommodation. The study contributes to research on Italian L2 and language contact by highlighting the role of regional phonetic patterns in shaping the acquisition of high-frequency pragmatic items.

Keywords: discourse markers; phonetic reduction; Italian speech; first-generation migrants; second-generation migrants; Filipino migrants; Sri Lankan migrants

1. Introduction

As is well-known, the Italian linguistic landscape is complex and highly heterogeneous, as it involves a number of linguistic varieties mainly differentiated along geographic and sociolinguistic axes (Berruto, 2012). In Italy, there is the standard language as it is taught in school, but it is not commonly spoken in everyday life except by trained speakers. Instead, Italian speakers are native speakers of varieties commonly referred to as ‘regional Italian(s)’, i.e., regionally marked varieties of spoken Italian (Cerruti, 2013; Telmon, 2016; Crocco, 2017). These varieties are reported to be influenced and shaped by substratum local dialects at multiple levels, including phonetics, morpho-syntax and lexicon. This is particularly evident in some Italian regions, such as Campania, whose regional variety of Italian appears to be easily recognizable to non-Campanians and coexists with a still very vital local dialect (Maturi, 2023), i.e., Neapolitan. The relationship between regional Italian and local dialect is characterized by the lack of clear-cut boundaries between such varieties, so that more ‘Italian’ traits can be used when speaking dialect as well as dialectal traits being able to interfere in (regional) Italian speech.



Academic Editor: John Hajek

Received: 30 March 2026

Revised: 12 May 2026

Accepted: 15 July 2026

Published: 31 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Such a sociolinguistic complexity appears to be even more complex for non-native speakers. Indeed, when migrants come to Naples, they are exposed to a range of intermediate spoken varieties that work as sources of linguistic input in their acquisition of Italian as a second or third/additional language. Furthermore, multilingual practices emerge among migrants, involving both varieties already included in their migratory repertoire and varieties that are present in the host environment.

These issues have been recently investigated within the HELLO Campania! Project (*Heritage Languages and languages of the Others in Campania*; <https://sites.google.com/view/prinpnrrhellocampania/home-page>, accessed on 10 May 2026), which focuses on the sociolinguistic study of different migrant communities (Sri Lanka, Ukraine, Bangladesh, the Philippines, Senegal, Ghana; Moro & Di Salvo, 2026). Within this framework, previous studies have examined linguistic contact and language choices among heritage language(s), regional Italian and Neapolitan dialect. Results show that the presence of local linguistic aspects—particularly at the phonetic and morpho-syntactic levels—covaries with a range of sociolinguistic variables, including migrant generation, ethnic group, speaker gender, exposure to local dialect and language attitudes (Cataldo, 2025; Di Salvo, 2025; Di Salvo et al., in press).

From a phonetic perspective, Sri Lankan and Ukrainian migrants living in Naples exhibit the presence of features shared by both regional Italian and local dialect, such as the palatalization of /s/ ([ʃ]) before consonants, the affrication of /s/ ([ts]) after nasals, diphthongs with mid–high vowels ([wo] and [je]), and the strengthening of /b/ and /dʒ/ in intervocalic position (Cataldo, 2025; Di Salvo & Cataldo, in press; Di Salvo et al., in press). These findings highlight the role of local input in shaping migrants' phonetic patterns.

Research on Italian L2 pragmatics has mainly dealt with the development of pragmatic competence, speech acts, and discourse markers involved in the management of textuality and interaction (Nuzzo & Santoro, 2017). Within this broader context, the study of discourse markers in the Italian speech of non-native speakers has mainly focused on functional aspects, i.e., occurrences of different discourse markers carrying out a set of pragmatic functions, and proficiency levels (Dal Negro & Fiorentini, 2014; De Cristofaro et al., 2024; De Marco & Paone, 2024). Indeed, following Carla Bazzanella's functional classification, functions performed by discourse markers can be ascribed to three main dimensions or macro-functions: interactional, metatextual, and cognitive (Bazzanella, 1995, 2006). Interactional functions concern the management of interpersonal relations and conversational dynamics, including turn-taking, checking comprehension, maintaining contact with the interlocutor, and ensuring communicative effectiveness. Metatextual functions involve the organization and structuring of discourse, contributing to textual cohesion and coherence through processes such as reformulation, clarification, exemplification, topic management, and planning. Cognitive functions, on the other hand, relate to the expression of inferential processes, speakers' attitudes, and the activation of shared knowledge and logical-argumentative relations. Importantly, discourse markers are highly multifunctional and may simultaneously operate across different levels of discourse and interaction. In the acquisitional literature, discourse markers in Italian L2 are reported to emerge following a functional hierarchy (Andorno, 2008), with interactional markers appearing earlier than metatextual and cognitive ones, reflecting learners' primary need to ensure communicative effectiveness in interaction.

Moreover, previous research has shown that discourse markers are particularly susceptible to language contact and cross-linguistic influence. For instance, Fiorentini (2016) demonstrates that reformulation markers such as *cioè* are frequently transferred in bilingual speech from the dominant language (Italian) into the minority language (Ladin), especially in core discourse-organizing functions such as paraphrase and correction. Crucially, the

study highlights that this transfer is not uniform across functions, but rather sensitive to the interactional and cognitive salience of specific marker types. However, while these studies provide important insights into the functional behavior of discourse markers in contact settings as well, detailed observations on their phonetic realization and potential phonetic reduction processes in L2 speech remain scarce.

Recently, [Cataldo \(in press\)](#) has shown that *cioè*, despite its high frequency and well-documented functional and positional profile, undergoes phonetic reduction in Campanian Italian, following a pathway from full to reduced forms, i.e., [tʃɔɛ] > [tʃɛ] > [ʃɛ]. This process reflects region-specific phonetic patterns—the lenition of /tʃ/ resulting in fricative [ʃ] ([Maturi, 2023](#))—and suggests that discourse markers may be particularly sensitive to local sociophonetic norms. Moreover, these findings shed light on regionally related reduction processes compared to previous research largely focused on northern or more standard-oriented Italian varieties ([Ghezzi, 2022](#); [Mereu & Dal Negro, 2025](#)).

The present study builds on this line of research by investigating the use of *cioè* in the Italian speech of first- and second-generation migrants settled in Naples. Among Italian discourse markers, *cioè* represents a particularly suitable candidate for investigating non-native speech. Indeed, it is very frequent in spoken Italian and, although it originally developed as a reformulation marker ([Bazzanella, 1995](#)), it has progressively expanded its functional range. In contemporary Italian, *cioè* can perform a variety of pragmatic functions beyond reformulation, including discourse organization, reinforcement and interaction management ([Dal Negro & Fiorentini, 2014](#); [Fiorentini & Sansò, 2017](#); [Sansò, 2020](#)). This functional expansion is accompanied by increased positional flexibility: indeed, it can occur in a final or isolated position as well ([Ghezzi, 2022](#)). From a phonetic perspective, *cioè* is typically produced at a high speech rate and is subject to phonetic reduction and hypo-specification ([Dal Negro & Fiorentini, 2014](#); [Geddo, 2023](#); [Barotto et al., 2024](#); [Mereu & Dal Negro, 2025](#); [Cataldo, in press](#)).

Accordingly, this study examines frequency, phonetic realization, pragmatic functions and syntactic position. It aims at providing an insight into migrants' linguistic practice of using a discourse marker within the framework of Campania by comparing their use with previous results concerning the Campanian area ([Cataldo, in press](#)). This fact plays a crucial role for this study, since it makes it possible to have a control group of native speakers without a migratory background to validate empirical hypotheses, as highlighted by [Di Salvo \(2025\)](#). In migrant contexts, where speakers are exposed to a range of input varieties, including regionally marked forms, such markers may be especially prone to both functional transfer and sociophonetic accommodation.

2. Materials and Methods

2.1. Participants

For this study, we relied on materials collected within the HELLO Campania! Project ([Moro et al., 2024](#); [Di Salvo et al., 2024](#)). In particular, structured sociolinguistic interviews were used, along the lines of previous studies exploring linguistic practices of migrants (see [Moro & Di Salvo, 2026](#)).

From among the various groups in HELLO Campania!, we selected the Sri Lankan (SL) and Filipino (FI) ethnic groups in order to have sufficiently large groups that are comparable between the first (G1) and second generations (G2) in terms of the number of speakers, especially regarding the second generation, which is small in other ethnic groups, the duration of the interviews, and the number of tokens. Details on the first- and second-generation datasets are provided in Tables 1 and 2, respectively.

Table 1. First generation dataset.

Ethnic group	Dataset—G1		
	n. speakers	Duration	n. token
SL	22	09:40:37	34.939
FI	14	10:48:26	27.678
tot	36	20:29:03	62.617

Table 2. Second generation dataset.

Ethnic group	Dataset—G2		
	n. speakers	Duration	n. token
SL	18	08:10:03	36.562
FI	17	11:36:00	37.286
tot	35	19:46:03	73.848

Both these ethnic groups have been previously described (see [Moro & Russo, 2025](#); [Cataldo, 2025](#); [Di Salvo, 2025](#)). The participant sample reflects a stratified migrant profile, with clear differences across ethnic groups and generations in terms of age, length of residence, socioeconomic integration, and linguistic competence. First-generation Sri Lankan speakers (n. 22) have a mean age of 35 and an average length of residence of 24 years in Italy. They display heterogeneous educational backgrounds and are primarily employed in domestic work. Previous studies have shown that this group typically exhibits basic or pre-basic proficiency in Italian ([Di Salvo et al., in press](#)), regardless of their length of residence and the reported attendance to Italian L2 classes. First-generation Filipino speakers (n. 14) are slightly older on average (mean age 41) and have resided in Italy for approximately 19 years. They are primarily employed in care work (e.g., domestic workers, caregivers) and in the service sector, which entails regular interaction with Italian speakers, although no detailed information on their level of Italian proficiency is currently available.

Second-generation Sri Lankan and Filipino speakers constitute comparable groups, with mean ages of 19 and 22, respectively. Both groups have been enrolled in or have completed the Italian schooling system, resulting in Italian-dominant linguistic repertoires and higher levels of fluency.

Clearly, both the two ethnic groups and the two generations differ as regards their linguistic repertoire. In particular, Table 3 shows that SL migrants homogeneously report to have Sinhala as their first language; on the other hand, FI migrants display broader multilingual repertoires including Filipino languages—primarily Tagalog—as well as regional languages. Previous studies revealed a maintenance-oriented practice toward heritage languages alongside functional multilingualism, depending on context, with English often used in community interaction and Italian in an institutional context, such as work and bureaucracy ([Moro & Russo, 2025](#)). Despite these differences, G1 participants show a similar average number of languages in their repertoire, with some also reporting competence in Neapolitan dialect.

Table 3. Linguistic repertoire of G1 participants.

Ethnic group	n. speakers	Repertoire—G1	
		First language	Average n. of languages
SL	22	Sinhalese (22)	4
FI	14	Tagalog (6); Ilokano (3); 2 Bisaya (2); Bikol (1); Cebuano (1); Aklanon (1)	5

Compared to the first generation, the repertoires of both Sri Lankan and Filipino second-generation participants are restructured. Table 4 shows G2 repertoires reporting both the first language that participants declared to have acquired during preschool age and their current dominant language. As can be noticed, although Sri Lankans declared themselves to have first acquired their heritage language more frequently than Filipinos, in both groups, the repertoires are characterized by Italian as the most frequently identified dominant language, with English as an additional resource, and limited or partial competence in heritage languages (Moro & Russo, 2025). At the same time, their repertoires remain complex and multilingual, with a higher average number of languages compared to G1, as all participants report competence in Italian, a heritage language and English, and approximately half also include Neapolitan.

Table 4. Linguistic repertoire of G2 participants.

Ethnic group	n. speakers	Repertoire—G2		Average n. of languages
		First language	Dominant language	
SL	18	Sinhala (12); Italian (3); English (2); Tamil (1)	Sinhala (4); Italian (9); English (4); Tamil (1)	4.5
FI	17	Italian (11); Tagalog (4); Ilokano (1); English (1)	Italian (13); Tagalog (3); English (2)	5.1

In addition to the migrant groups described above, the present study also draws on the results of a previous investigation focusing on the use of *cioè* in Campanian regional Italian (Cataldo, in press). That study included two groups of native speakers from Naples, differentiated by age, which serve as a baseline for native (local) usage. This control dataset (NA; Table 5) allows for a systematic comparison with first- and second-generation migrant speakers, providing a reference point for interpreting patterns of frequency, phonetic realization, pragmatic function, and syntactic position of *cioè*. In particular, the study relied on 40 free-topic conversations (about 17 h of speech) by couples of speakers from Naples. The dataset was part of a larger corpus under construction collecting spoken Italian of Neapolitans (Di Salvo, in press). The sample consists of two groups of speakers differentiated according to age: 20 university students between 21 and 29 years old (9 M, 11 F) and 20 adults with 63 years as average age (2 M, 18 F).

Table 5. NA control group dataset.

Dataset—NA			
Group	n. speakers	Duration	n. token
Young NA	20	08:28:59	58.307
Adult NA	20	08:09:53	51.093
tot	40	16:38:52	109.400

2.2. Data Annotation

Interviews had already been transcribed and tokenized (see Cataldo, 2025). For the aims of this study, further multi-level annotation in Elan was carried out (Sloetjes & Wittenburg, 2008) in order to annotate discourse markers across different levels. Specifically, to ensure comparability with previous research, the annotation followed the workflow employed by Mereu and Dal Negro (2025) and employed in Cataldo (in press) for discourse markers labeling in Campanian Italian speech. Each occurrence of *cioè* was therefore annotated on the following levels.

1. Pragmatic function performed by *cioè* within the discourse according to the three-way dimensions identified by Bazzanella (1995). Below are examples (1–7) for each function, taken from interviews with the participants.
 - Turn-taking (TURN_T) function within the interactional dimension; in the following example (1), the speaker interrupts the interviewer's turn and takes the floor by means of *cioè*.
 - (1) Interviewer: napoletano inglese e italiano?
'Neapolitan English and Italian?'
FI2F21: sì <sp> sì <sp> è sempre un misto il mio parlare
'yes <sp> yes <sp> my speech is always a mixture'
Interviewer: quindi veramente le mischi tutte queste lingue
'so you really mix all these languages together'
FI2F21: *cioè* io faccio *il switch <sp> se io ti parlo in italiano <sp> belle e buono me viene in inglese <sp> così come può venire in tagalog
'**I mean** I switch languages <sp> if I'm speaking to you in Italian <sp> at some point I might switch to English <sp> just as I might switch to Tagalog'.
 - Paraphrastic reformulation (PAR_REF) within the metatextual dimension; in example (2), the speaker reformulates a segment ('treated') with a semantically equivalent segment ('treated').
 - (2) SL1M12: questo sì succede <eeh> dappertutto in ogni paese sì può capitare anche se un italiano va in mio paese <sp> si può essere novanta percento sì <eeh> tratta *cioè* tratta be+ benissimo <eeh> qualcuno raramente dice qualcosa si può capitare per ognuno non dipende da paese o<oo>
'this yes happens <eeh> everywhere in every country yes it can happen even if an Italian goes to my country <sp> it can be ninety percent is <eeh> treated **that is** treated we+ very well <eeh> anyone rarely says anything it can happen to anyone it doesn't depend on the country or<rr>'.
 - Non-paraphrastic reformulation (NON_PAR_REF) within the metatextual dimension; in this case (3), *cioè* is used to reformulate a segment ('I'm used to speaking Italian') into a non-semantically equivalent one ('since I'm here', i.e., in Italy).
 - (3) FI2F12: allora l'italiano perché comunque appunto essendo abituata a parlare in italiano <sp> *cioè* stando qui <sp> è la lingua principale che parlo.
'so, Italian because anyway exactly I'm used to speaking Italian <sp> **that is** since I'm here <sp> it's the main language I speak'.
 - Planning device (PLAN), within the metatextual dimension. In example (4), *cioè* is employed as a way to take time to plan upcoming speech; indeed, it co-occurs with other filler and/or hesitation phenomena, such as silent or filled pauses and segmental prolongation (*cioè<ee>*).

- (4) Interviewer: *ok <sp> secondo te c'è discriminazione contro le persone di origine filippina?*
 'okay <sp> do you think there's discrimination against people of Filipino descent?'
 FI2F05: *dipende se in positivo o negativo*
 'it depends on whether it's positive or negative'
 Interviewer: *discriminazione in negativo*
 'negative discrimination'
 FI2F05: *<ehm> non proprio<oo> <sp> cioè<ee> <sp> secondo me dipende insomma da dove ti trovi perché in realtà vengono viste molto bene le persone filippine*
 '<ehm> not really<yy> <sp> **I mean**<nn> <sp> I think it depends on where you are because Filipino people are actually viewed very favorably'.
- Demarcation (DM) function within the metatextual dimension; in example (5), *cioè* introduces a parenthetical clause ('we take part of that festival'), which adds extra information to the main clause.

(5) SL2M10: *<eeh> ogni anno io vado sempre a Sri Lanka perché prima di tutto <eeh> là tenevo al<ll> a un tempio <sp> una festività in cui ogni anno cioè noi facciamo parte di quella festività quindi ogni anno tipo i momenti di febbraio andiamo sempre a Sri Lanka <sp> per un mese*
 '<eeh> every year I always go to Sri Lanka because first of all <eeh> there I used to have at<tt> at a temple <sp> a festival in which every year **I mean** we take part of that festival so every year around February we always go to Sri Lanka <sp> for a month'.
 - Reinforcement (BOOST), within the cognitive dimension, with *cioè* working as an intensifier of the illocutionary force of the speech act (example (6)).

(6) Interviewer: *<ehm> <sp> i tuoi genitori ti hanno mai detto di come sono stati accolti<ii> quando sono arrivati in Italia sono stati accolti bene?*
 '<ehm> <sp> did your parents ever tell you how they were received<dd> when they arrived in Italy were they welcomed?'
 SL2F04: *vabe le difficoltà che hanno avuto sì le hanno<oo> raccontate <sp> cioè comunque hanno avuto molte difficoltà perchè comunque rispetto al loro livello di studio e i lavori che fanno <sp> <eeh> c'è un divario enorme*
 'well the difficulties they had yes they did tell about it <sp> **I mean** anyway they had a lot of difficulties because considering their level of education and the jobs they do <sp> <eeh> there is a huge gap'.
 - Mitigation (HEDGE), within the cognitive dimension. In example (7), *cioè* is used to mitigate the rather strong effect of the preceding expression.

(7) Interviewer: *ma l'inglese anche con gli amici <sp> per caso ti è capitato di usarlo?*

'but English even with friends just by chance have you ever found yourself speaking it?'

FI2F12: *a parlarlo frequentemente <sp> no <sp> cioè <sp> comunque magari più che altro a scuola o durante magari i corsi extracurricolari*
'I don't really speak it that often <sp> no <sp> **that is** <sp> maybe mostly at school or during extracurricular activities'.

2. Syntactic positions occupied by *cioè* within the dominion of the slash-units, suitable for conversational speech accounting for both syntactic and intonational criteria (Meteer et al., 1995; Bell et al., 2003; Mereu & Dal Negro, 2025):
 - Initial;
 - Medial;
 - Final;
 - Stand-alone.
3. Phonetic realization: the annotation was then imported into Praat (Boersma & Weenink, 2022) to carry out acoustic and spectral analysis and identify phonetic forms.

Figure 1 shows an example of multi-level annotation.



Figure 1. Multi-level annotation in Elan. Tiers from top to bottom: speaker 1 (turn of the interviewer); speaker 2 (turn of the sample speaker); token (tokenization of speaker 2); KP_DM_function (discourse marker pragmatic function); KP_DM_position (discourse marker syntactic position); KP_DM_phonetic (discourse marker phonetic transcription).

2.3. Statistical Analysis

The data were statistically analyzed in R version 4.0.1 (R Core Team, 2020). The analysis considered both extralinguistic and linguistic factors. Speaker group and gender were treated as independent (extralinguistic) variables, while phonetic form, pragmatic function and syntactic position were modeled as dependent (linguistic) variables.

First, contingency analyses were conducted using chi-square tests, followed by the inspection of Pearson residuals to identify significant deviations from expected frequencies. Standardized Pearson residuals were interpreted using a threshold of ± 2 . Pairwise comparisons of proportions were carried out using proportion tests (prop.test).

In addition, the data were modeled using both simple and mixed-effects generalized linear models (GLM and GLMM), implemented through the *lme4* package (Bates et al., 2015). Multinomial logistic regression models were fitted using the *nnet* package (Venables & Ripley, 2002), in order to account for categorical dependent variables with more than two levels. In mixed-effects models, speaker was included as a random effect to control for inter-speaker variability.

3. Results

Results of the study are reported for the first generation (§ 3.1) and second generation (§ 3.2). As already said, findings related to migrants' speech are compared with previous observations on native speakers of Campanian Italian. Specifically, as Cataldo (in press) points out, young and adult native Campanians differ according to age with respect to the frequency of use of *cioè*, with a higher frequency in young sample. Overall, Campanian Italian, along the same lines as Mereu and Dal Negro (2025) with respect to Bolzano Italian, exhibit a neutral distribution of the sample with the most frequent phonetic form [tʃe] occurring in medial position and mostly performing paraphrastic and non-paraphrastic reformulation; in Campanian Italian, the second and more reduced forms [ʃe] tend to occur in medial position and are associated with demarcation and mitigation functions.

3.1. First Generation

In the G1 dataset, 74 occurrences of *cioè* were detected (Table 6). Specifically, these few instances (26 in SL and 48 in FI) indicate a very low incidence of *cioè* per 1.000 words (0.7 and 1.7, respectively). If, on the one hand, this finding is in line with the low frequency of *cioè* in adult native speakers of Naples as well (1.6), the present occurrences can only be discussed qualitatively, as they were found in the speech of few G1 speakers in both the SL and FI groups.

Table 6. *Cioè* in G1 dataset.

Ethnic Group	Occ. <i>Cioè</i>	Incidence
SL	26	0.7
FI	48	1.7
tot	74	

As regards the phonetic forms in which *cioè* may occur, FI shows a distribution that, again, is comparable to native adult Campanians, i.e., an almost exclusive use of the [tʃe] variant (92%). SL, instead, presents [tʃe] as the majority variant (65%), alternating with the minority locally marked variant [ʃe] (31%). It should be noted that all instances of the minority variant were produced by a single speaker (SL1M12), who presents a sociolinguistic profile that makes him prone to a more 'Campanian' realization of *cioè*: a 40-year-old man who is employed as a domestic worker, attends Italian language classes and is exposed to Neapolitan, although he is not the only one within SL G1.

As for pragmatic functions, SL and FI exhibit a similar distribution: *cioè* mostly performs functions of non-paraphrastic reformulation (34% and 44%) and boost (19% in both groups), with all the other functions being underrepresented, except for planning function, which presents a much higher percentage of use with respect to Campanian speakers: 17% in FI and 35% in SL, being the most common function performed by *cioè* in migrants' speech. While this finding differs from the Campanian native distribution, it also shows that first-generation migrants with low levels of Italian proficiency do indeed need more time to plan their speech in the second language, but are also aware of the possibility of using *cioè* as a means of taking and managing time in the construction of their speech, hence as 'hesitation markers' (Andorno, 2008).

The occurrence of *cioè* according to syntactic functions is in line with the most frequent functions. As expected, *cioè* primarily occurs in medial position in both SL and FI groups (85% and 63%, respectively) as in the control group, and is associated with the function of reformulation, with an interesting and higher percentage of initial position (SL: 11%; FI: 29%) which, instead, tends to co-occur with planning function. Overall, the low fre-

quency of *cioè* in G1 suggests limited integration of discourse-pragmatic devices at early stages of L2 acquisition.

3.2. Second Generation

Turning to the second generation (Table 7), a much higher number of occurrences was identified—678 in total. The incidence per 1.000 words is very different from that of the first generation and very similar to that of the group of young Neapolitans (8.5).

Table 7. *Cioè* in G2 dataset.

Ethnic Group	Occ. <i>Cioè</i>	Incidence
SL	376	10.3
FI	302	8.1
tot	678	

3.2.1. Phonetic Form

In order to explore the phonic forms employed by SL and FI second-generation speakers, Figure 2 shows the distribution of usage percentages of *cioè* in descending order of frequency, including the control group of young native speakers of Naples (NA).

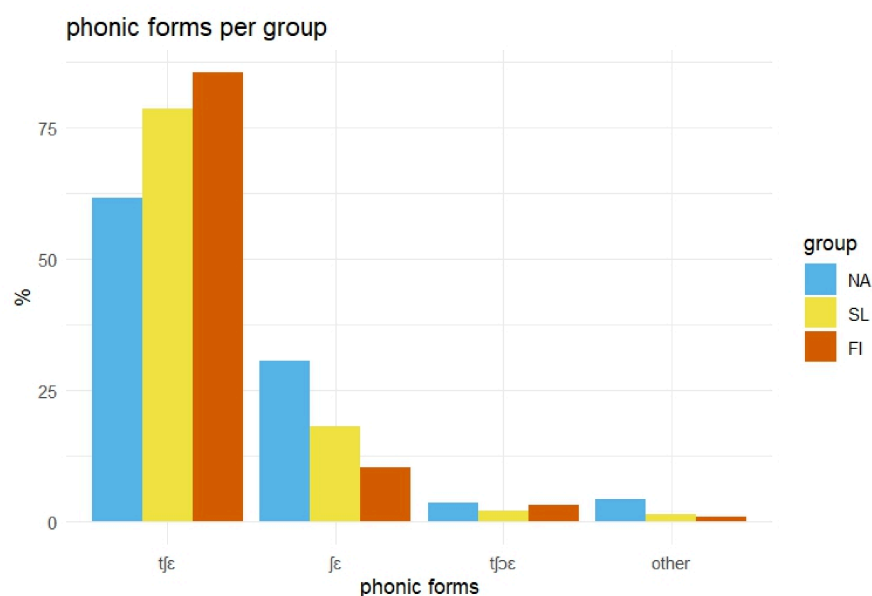


Figure 2. Distribution of phonetic forms per group (NA, SL, FI).

As the bar chart shows, the full bisyllabic form, which is extremely rare in NA, is equally rare in SL and FI (around 3% of occurrences). The most frequent form by far is [tʃɛ], which, as already mentioned, is the default form of *cioè* (Mereu & Dal Negro, 2025; Cataldo, *in press*). This is true for NA and even more so for SL, who use it in 78% of cases, and for FI in 85% of cases. In all three groups, the further reduced form [ʃɛ] works as a minority variant. While NA uses this variant one-third of the time (30%), second-generation speakers use it less frequently (SL: 18%; FI 10%). The ‘other’ category includes other possible but occasional forms, e.g., [dʒɛ], [ʃə], [tʃɛ], for a total of 60 other occurrences.

Statistical analysis was carried out by taking into account the selection of the two most frequent variants ([tʃɛ], [ʃɛ]) and the three groups of speakers (NA, SL, FI) in order to explore whether the distribution of phonic forms across the speaker groups was statistically different. The chi-square test was highly significant (X-squared = 55.577, df = 2, *p*-value < 0.001), showing a significant difference between the groups, as well as

the prop.test used for performing pairwise comparisons that highlighted significant differences in the frequency of use of [tʃɛ] both between NA and SL (X-squared = 20.786, df = 1, p -value < 0.001) and NA and FI (X-squared = 47.043, df = 1, p -value < 0.001). Table 8 summarizes these results, highlighting the significantly higher percentage of use of [tʃɛ] in SL and FI with respect to NA.

Table 8. Comparison in the use of [tʃɛ] across the three groups.

Ethnic Group	% [tʃɛ]	Difference vs. NA	p -Value
SL	81.3%	+14.5%	<0.001
FI	89.3%	+22.5%	<0.001
NA	66.8%	–	–

Since in NA, speaker gender had an effect on the selection of the two variants, with male speakers more often preferring the reduced form [ʃɛ], for the present analysis, the gender variable in G2 speakers was added. The distribution of the two phonic variants according to group and gender (M, F) is shown in Figure 3.

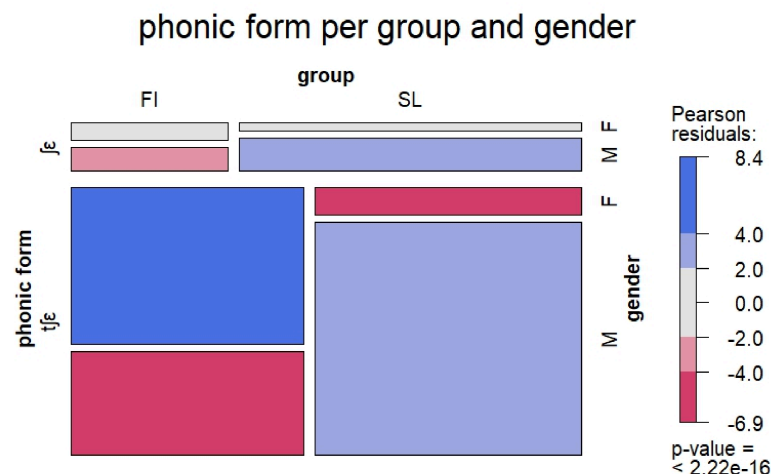


Figure 3. Mosaic plot of phonic form per speaker group and gender.

The distribution shown in the mosaic plot is confirmed by a highly significant chi-square test revealing a significant relationship between phonic form and the group-gender classification (X-squared = 61.954, df = 5, p -value < 0.001). Pearson's residuals show that female FIs use the form [ʃɛ] less (residual = −4.05) and [tʃɛ] much more (residual = +2.19) and male SLs use [ʃɛ] less than expected (−2.01). In summary, female FIs and male SLs use [ʃɛ] less than expected with respect to the control group NA.

3.2.2. Pragmatic Functions

As for pragmatic functions performed by *cioè*, the bar plot in Figure 4 shows the percentage distribution across the three groups (SL, FI, NA). As can be seen, *cioè* is used with a similar pragmatic distribution across the three groups, i.e., in G2 SL and FI with respect to NA.

Specifically, NON_PAR_REF appears to be the most frequent function, followed by BOOST; other functions are similarly distributed, while DM and TURN_T are underrepresented. No gender-related effects were found, while the chi-square test considering the three groups resulted in being significant (X-squared = 37.639, df = 12, p -value < 0.001). The analysis of Pearson residuals shows that such differences regard group-specific prefer-

ences: FI speakers use the BOOST function less (−2.91) and SL speakers use PAR_REF less compared to NA.

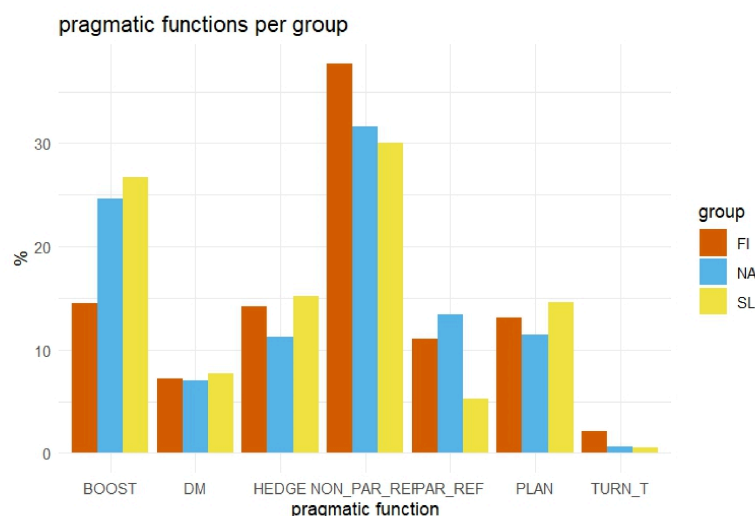


Figure 4. Distribution of pragmatic functions per group (NA, SL, FI).

3.2.3. Syntactic Positions

Second-generation and control group distribution of syntactic positions occupied by *cioè* is shown in Figure 5. As expected, second-generation groups use *cioè* in medial position, in line with native speakers. Distribution related to other positions (initial, final, stand-alone) shows a similar tendency as well. Significant differences emerge (X-squared = 24.886, $df = 6$, p -value < 0.001), revealing that FIs use *cioè* in initial and final position more often than both SLs and NAs (2.57 and 2.25, respectively), while SLs use it less often in initial position (−2.36).

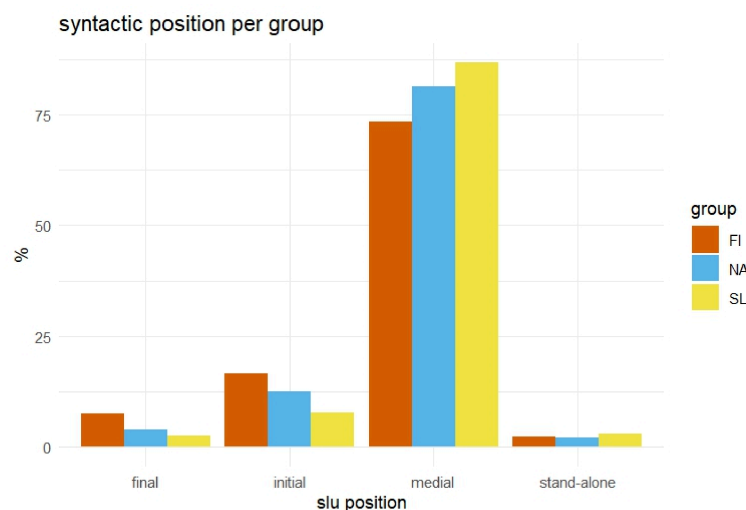


Figure 5. Distribution of syntactic positions per group (NA, SL, FI).

4. Discussion

The present study aimed at exploring the use of the discourse marker *cioè* in the Italian speech of migrants living in Naples belonging to two different ethnic groups and migrant generations. The investigations aimed to shed light on the use of *cioè* in terms of frequency, phonic realization, pragmatic functions and syntactic positions, building on recent studies

that provide a detailed account of native usage of this discourse marker with a specific look at its phonetic regional variation (Mereu & Dal Negro, 2025; Cataldo, *in press*).

Overall, results confirm that *cioè* is a highly frequent discourse marker, whose distribution in migrant speech closely mirrors that of native speakers. In particular, its frequency is higher among young Neapolitans and second-generation migrants than among adult Neapolitans and first-generation migrants. At the same time, migrant speakers not only acquire *cioè* at a functional level, but also reproduce phonetic realizations that are typical of Campanian regional Italian. Specifically, across both ethnic groups and generations, [tʃɛ] emerges as the unmarked—default—variant, as pointed out in recent studies (Barotto et al., 2024; Mereu & Dal Negro, 2025). However, important generational differences arise. In first-generation speakers, [tʃɛ] is the only attested variant, whereas in second-generation speakers, it alternates with the reduced variant [ʃɛ], characterized by phonetic reduction and the lenition of /tʃ/ into [ʃ]. Crucially, this alternation closely reflects the distribution observed in native Campanian speech. Pragmatic functions and syntactic position also show convergence with native usage, with *cioè* predominantly used for paraphrastic and non-paraphrastic reformulation and reinforcement, and typically occurring in medial position. These patterns largely replicate those observed in the control group (Cataldo, *in press*).

Taken together, these findings point to a process of sociolinguistic accommodation that extends to high-frequency pragmatic items. Migrant speakers not only adopt *cioè* as a functional discourse marker (Fiorentini, 2016), but also converge toward locally salient phonetic variants, including reduced forms such as [ʃɛ]. This suggests that the integration of discourse markers in contact situations operates across multiple levels, involving both functional transfer and phonetic accommodation.

The results for first-generation speakers show a low frequency of *cioè* and a more limited functional range. This can be related to lower proficiency in Italian, as also reported in previous studies (Giuliano & Russo, 2014; Giuliano et al., 2019; De Marco & Paone, 2024). In particular, research has shown that a long—if not very long—length of residence in Italy, especially among Sri Lankan migrants, does not necessarily lead to high levels of proficiency in Italian. In line with this, first-generation speakers tend to rely more heavily on interactional discourse functions, aimed at ensuring communicative effectiveness, rather than on metatextual functions such as reformulation, which constitutes the core function of *cioè* (Andorno, 2008). This helps explain both the lower frequency and the less specialized use of the marker in this group.

In contrast, second-generation speakers display a much higher frequency of *cioè* and a wider, more native-like functional range. Their usage closely aligns with that of native speakers, not only in terms of distribution across pragmatic functions and syntactic positions, but also in terms of phonetic variation. The emergence of the reduced variant [ʃɛ], alongside the default [tʃɛ], indicates an advanced level of sociophonetic convergence with local norms. As described in previous works (Moro & Russo, 2025), second-generation migrants—due to their greater exposure to local Italian—exhibit a shift in actual practices toward an increasing dominance of Italian. This study highlights that these speakers, having been socialized within the Italian linguistic environment, are more sensitive to locally salient phonetic patterns and are able to incorporate them into their speech.

These findings contribute to a broader understanding of linguistic practices in migrant contexts. Previous research has identified region-specific phonetic patterns in Campanian Italian, including a reduction pathway for *cioè* from /tʃɔɛ/ to /tʃɛ/ and further to /ʃɛ/ (Cataldo, *in press*), which differs from patterns observed in more standard-oriented varieties such as Bolzano Italian (Mereu & Dal Negro, 2025). The present study shows that these locally grounded patterns are part of the input available to migrant speakers—in the

workplace, during Italian classes—and can be reproduced in their speech, regardless of their degree of exposure to dialect.

At the same time, unlike other phonetic features more strictly intertwined with the degree of exposure of migrants to linguistic varieties closer to the dialect and therefore to dialect contact (e.g., /s/ palatalization; Cataldo, 2025), the reduced forms of *cioè* appear to be widespread across different interactional contexts, making them particularly accessible to learners. In this respect, ethnic group- and gender-related differences emerging in previous research do not substantially differentiate migrants taken into consideration here, whereas generation emerges as the key factor shaping linguistic behavior.

More generally, these results support a model of linguistic integration as a multi-layered process, in which different components of language are acquired through distinct mechanisms and at different rates. While morpho-syntactic convergence may remain limited (Di Salvo, 2025), speakers can exhibit strong alignment with local norms at the phonetic and pragmatic levels. This points to a level-sensitive model of linguistic integration, in which different components of language—morpho-syntax, pragmatics, and phonetics—follow different developmental trajectories and are shaped by different types of input and social interaction.

Despite the insights offered by this study, a number of limitations need to be considered. First, no detailed information is currently available on the level of Italian proficiency of the Filipino group, which may limit the interpretation of some of the observed patterns, particularly in comparison with the Sri Lankan speakers. Moreover, the comparison between first-generation migrants and adult Neapolitan speakers is intended primarily as a general baseline rather than a fully balanced comparison, due to differences in terms of average age and native vs. non-native speaker status. On the other hand, the inclusion of a native control group makes it possible to observe patterns of convergence and divergence in the linguistic practices of second-generation migrant speakers, providing clearer evidence of alignment with—or deviation from—native usage (Di Salvo, 2025).

5. Conclusions

The present study has explored the use of the discourse marker *cioè* in the Italian speech of Sri Lankan and Filipino migrants living in Naples, comparing first- and second-generation speakers with a native Campanian control group. By examining the frequency of use, pragmatic functions, syntactic position and phonetic realization, the study provides a multidimensional account of how discourse markers behave in migrant speech and in situations of language contact.

The analysis revealed clear generational differences. First-generation speakers displayed a low frequency of use and a more restricted functional range, whereas second-generation speakers exhibited more Campanian native-like distributions in terms of both pragmatic and positional patterns of *cioè* and local phonetic variants, including the reduced form [ʃɛ].

These findings suggest that the acquisition of discourse markers in migrant contexts involves both pragmatic competence and sociophonetic accommodation. By focusing on a highly frequent and multifunctional discourse marker, this study contributes to research on Italian L2/L3, discourse markers and language contact, highlighting the role of regional input in shaping migrants' linguistic practices.

Funding: This research was funded by the PRIN PNRR Project HELLO Campania (Heritage Languages and Languages of the Others, #P2022WJ8YF).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of University of Naples Federico II (protocol code PG/2024/0082451 and date of approval 28 June 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Research data presented in this study are available in the Eurac Research CLARIN Center. Sri Lanka: [Di Salvo et al. \(2024\)](#). The Philippines: [Moro et al. \(2024\)](#).

Acknowledgments: The author would like to thank Margherita Di Salvo for her encouragement and constant support for this work.

Conflicts of Interest: The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

SL	Sri Lankan
FI	Filipino
NA	Neapolitan
G1	First generation
G2	Second generation
TURN_T	Turn-taking (function)
PAR_REF	Paraphrastic reformulation (function)
NON_PAR_REF	Non-paraphrastic reformulation (function)
PLAN	Planning device (function)
DM	Demarcation (function)
BOOST	Reinforcement (function)
HEDGE	Mitigation (function)

References

- Andorno, C. (2008). Connettivi in italiano L2 fra struttura dell'enunciato e struttura dell'interazione. In G. Bernini, L. Spreafico, & A. Valentini (Eds.), *Competenze lessicali e discorsive nell'acquisizione di lingue seconde* (pp. 481–510). Guerra.
- Barotto, A., Combei, R., Fiorentini, I., & Sansò, A. (2024, September 19–21). Cioè/cè: *Phonetic and prosodic features in the pragmatic developments of an Italian reformulation marker*. “Varietà parlate, segnate e scritte: Nuovi paradigmi interpretativi tra sincronia e diacronia”—LVII Congresso Internazionale Della Società Di Linguistica Italiana, Catania, Italy.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. [[CrossRef](#)]
- Bazzanella, C. (1995). I segnali discorsivi. In L. Renzi, G. Salvi, & A. Cardinaletti (Eds.), *Grande grammatica italiana di consultazione* (Vol. 3, pp. 225–257). Il Mulino.
- Bazzanella, C. (2006). Discourse markers in Italian: Towards a “compositional” meaning. In K. Fischer (Ed.), *Approaches to discourse particles* (pp. 449–464). Elsevier.
- Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, 113(2), 1001–1024. [[CrossRef](#)] [[PubMed](#)]
- Berruto, G. (2012). *Sociolinguistica dell'italiano contemporaneo*. Carocci.
- Boersma, P., & Weenink, D. (2022). *PRAAT: Doing phonetics by computer* (Version 6.2.06) [Computer program]. Available online: <http://www.praat.org> (accessed on 28 September 2025).
- Cataldo, V. (2025). Phonetic features of Neapolitan and Campania regional Italian in migrants' speech. *International Journal of Linguistics*, 17(5), 201–220. [[CrossRef](#)]
- Cataldo, V. (in press). Discourse markers between phonetic reduction and regional Italian.
- Cerruti, M. (2013). Varietà dell'italiano. In G. Iannàccaro (Ed.), *La linguistica italiana all'alba del terzo millennio (1997–2010)* (pp. 91–127). Bulzoni.
- Crocco, C. (2017). Everyone has an accent. In M. Cerruti, C. Crocco, & S. Marzo (Eds.), *Towards a new standard: Theoretical and empirical studies on the restandardization of Italian* (pp. 109–128). De Gruyter.

- Dal Negro, S., & Fiorentini, I. (2014). Reformulation in bilingual speech: Italian *cioè* in German and Ladin. *Journal of Pragmatics*, 74, 94–108. [CrossRef]
- De Cristofaro, E., Badan, L., & Belletti, A. (2024). Discourse markers in L1 and L2 Italian: A cartographic analysis of the sentence-internal position. *Second Language Research*, 40(3), 643–673.
- De Marco, A., & Paone, E. (2024). Exploring discourse markers in complaints: A comparative analysis of native and non-native Italian speakers. *Italiano LinguaDue*, 16(2), 122–141.
- Di Salvo, M. (2025). Italiano regionale e campano come L2 di srilankesi di prima e seconda generazione: Un quadro sociolinguistico. *Italiano LinguaDue*, 17(1), 614–631. [CrossRef]
- Di Salvo, M. (in press). *Modulo ParlaNA nel corpus KIParla*.
- Di Salvo, M., & Cataldo, V. (in press). La comunità ucraina a Napoli tra italiano e dialetto: Variazione di genere e napoletanità. *Rassegna italiana di linguistica applicata*.
- Di Salvo, M., Cataldo, V., & Maffia, M. (in press). Pure Italian and impure dialect: Attitudes and uses in two groups of immigrants in Naples (Italy). *Analisi Linguistica e Letteraria*.
- Di Salvo, M., Cataldo, V., Maffia, M., & Noschese, M. P. (2024). *HELLO CAMPANIA! Sri Lanka collection*. Eurac Research CLARIN Centre. Available online: <https://hdl.handle.net/20.500.12124/87> (accessed on 19 February 2026).
- Fiorentini, I. (2016). Segnali discorsivi italiani in situazione di contatto linguistico: Il caso degli indicatori di riformulazione. *Quaderns d'Italià*, 21, 0011–0026.
- Fiorentini, I., & Sansò, A. (2017). Reformulation markers and their functions: Two case studies from Italian. *Journal of Pragmatics*, 120, 54–72. [CrossRef]
- Geddo, C. (2023). Parafrasare, ripetere, riformulare: Il caso di *cioè/cè* nel parlato contemporaneo. In D. Mastrantonio, I. G. M. Abdelsayed, M. Marrucci, M. Bellinzona, O. Paris, & V. Bianchi (Eds.), *Repetita iuvant, perseverare diabolicum. Un approccio multidisciplinare alla ripetizione* (pp. 57–69). Edizioni Università per Stranieri di Siena.
- Ghezzi, C. (2022). *Vagueness Markers in Italian: Age variation and pragmatic change*. FrancoAngeli.
- Giuliano, P., & Russo, R. (2014). L'uso dei marcatori discorsivi come segnale di integrazione linguistica e sociale. In P. Donadio, G. Gabrielli, & M. Massari (Eds.), *Uno come te. Europei e nuovi europei nei percorsi di integrazione* (pp. 237–247). FrancoAngeli.
- Giuliano, P., Russo, R., & Anastasio, S. (2019). Discourse markers, interlanguage level and social integration: The immigrant learners in the Naples area. In *Cognitive insights into discourse markers and second language acquisition* (pp. 227–248). Peter Lang.
- Maturi, P. (2023). *Napoli e la campania*. Il Mulino.
- Mereu, D., & Dal Negro, S. (2025). Pragmatic functions and phonetic reduction: *Cioè* and *cè* in contemporary spoken Italian. *Journal of Pragmatics*, 236, 1–14. [CrossRef]
- Meteer, M., Taylor, A. A., MacIntyre, R., & Iyer, R. (1995). *Dysfluency annotation stylebook for the switchboard corpus*. University of Pennsylvania.
- Moro, F. R., & Di Salvo, M. (2026). The 'HELLO Campania' project and corpus: Documenting language practices of linguistic minorities in Campania, Italy. *ETHNORÊMA*, 22, 1–38.
- Moro, F. R., & Russo, G. (2025). Family language policy in multilingual Filipino families in Italy. *Journal of Multilingual and Multicultural Development*, 46(9), 2965–2979. [CrossRef]
- Moro, F. R., Russo, G., & Corotan, M. (2024). *HELLO CAMPANIA! Philippines collection*. Eurac Research CLARIN Centre. Available online: <https://hdl.handle.net/20.500.12124/89> (accessed on 19 February 2026).
- Nuzzo, E., & Santoro, E. (2017). Apprendimento, insegnamento e uso di competenze pragmatiche in Italiano L2/LS: La ricerca a partire dagli anni Duemila. *EuroAmerican Journal of Applied Linguistics and Languages*, 4(2), 1–27. [CrossRef]
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Sansò, A. (2020). *I segnali discorsivi*. Carocci.
- Sloetjes, H., & Wittenburg, P. (2008, May 26–June 1). *Annotation by category-ELAN and ISO DCR*. 6th International Conference on Language Resources and Evaluation (LREC 2008) (pp. 816–820), Marrakech, Morocco.
- Telmon, T. (2016). Gli italiani regionali. In S. Lubello (Ed.), *Manuale di linguistica italiana* (pp. 301–327). De Gruyter.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. Springer.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.