



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Il corpus KIParla

Caterina Mauri

13/5/2025

Language tips, Università del Piemonte Orientale

** La ricerca che ha portato a questi risultati ha ricevuto il finanziamento del Progetto "DiverSIta-Diversità nell'italiano parlato", prot. P2022RFR8T, CUP J53D23017320001, finanziato dall'Unione Europea nel piano NextGenerationEU attraverso il "Bando Prin 2022 - D.D. 1409 del 14-09-2022".*

Panoramica

1. Progettazione e implementazione del corpus

2. I moduli

3. Utilizzo del corpus: prospettive ed esempi



Panoramica

1. Progettazione e implementazione del corpus

2. I moduli

3. Utilizzo del corpus: prospettive ed esempi



Progettazione del corpus KIParla ~~impresa!~~

Il corpus KIParla si propone di offrire:

- ✓ corpus **liberamente accessibile**

trascrizioni allineate ai file audio

- ✓ sistema di **metadati trasparente**

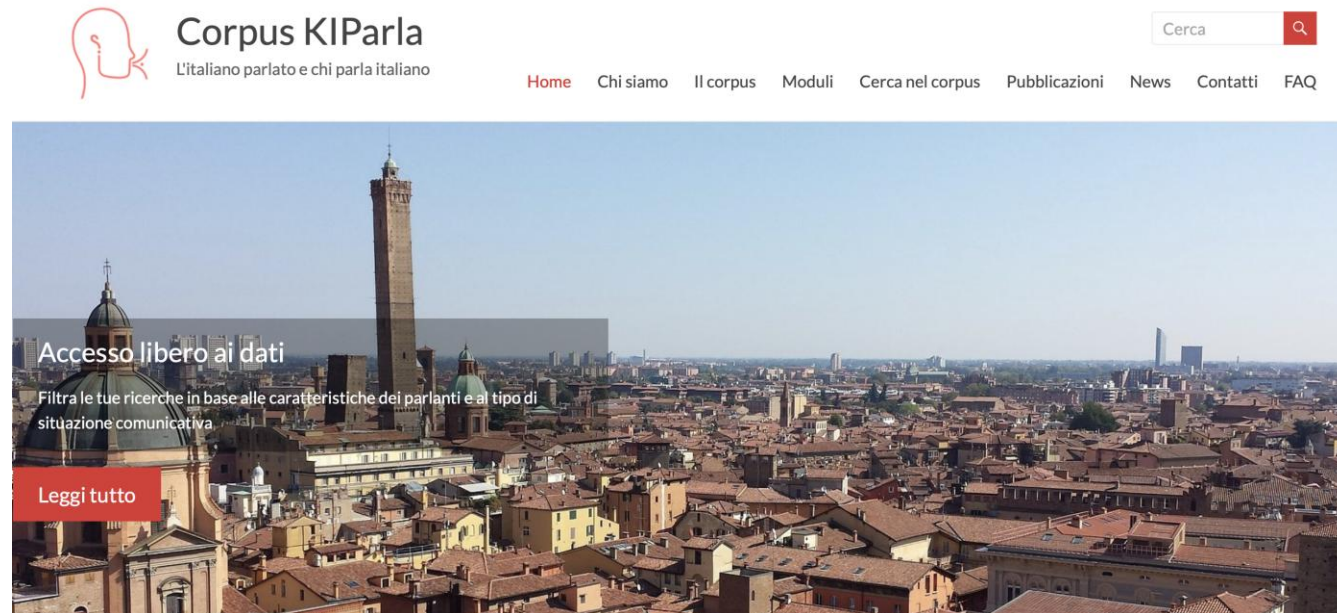
sia per quanto riguarda le situazioni comunicative sia per quanto riguarda i parlanti coinvolti

- ✓ **interfaccia di ricerca** basata su uno **standard internazionale**

NoSketchEngine, che offre funzioni di ricerca avanzate: KWIC, elenchi di frequenza, ecc.

- ✓ **infrastruttura modulare, incrementale e replicabile**

consentire l'espansione del corpus nel tempo attraverso una struttura modulare



www.kiparla.it

(Mauri et al. 2019)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Progettazione del corpus KIParla **impresa!**

Il corpus KIParla si propone di offrire:

- ✓ corpus **liberamente accessibile**

trascrizioni allineate ai file audio

- ✓ sistema di **metadati trasparente**

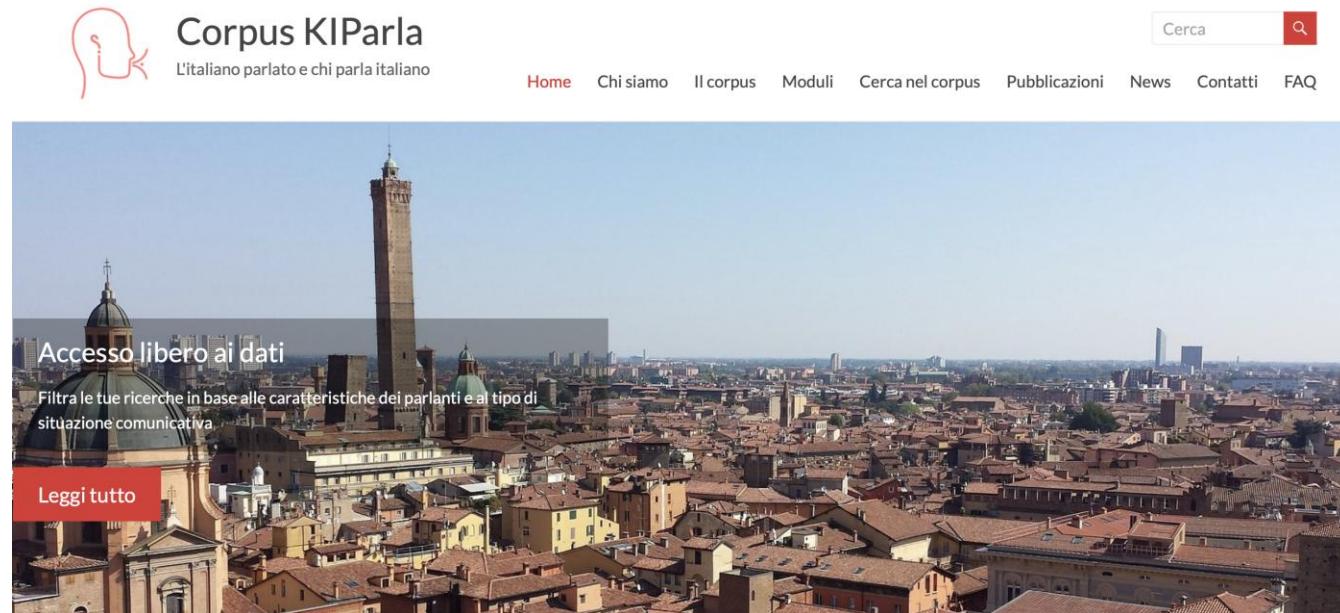
sia per quanto riguarda le situazioni comunicative sia per quanto riguarda i parlanti coinvolti

- ✓ **interfaccia di ricerca** basata su uno **standard internazionale**

NoSketchEngine, che offre funzioni di ricerca avanzate: KWIC, elenchi di frequenza, ecc.

- ✓ **infrastruttura modulare, incrementale e replicabile**

consentire l'espansione del corpus nel tempo attraverso una struttura modulare



www.kiparla.it

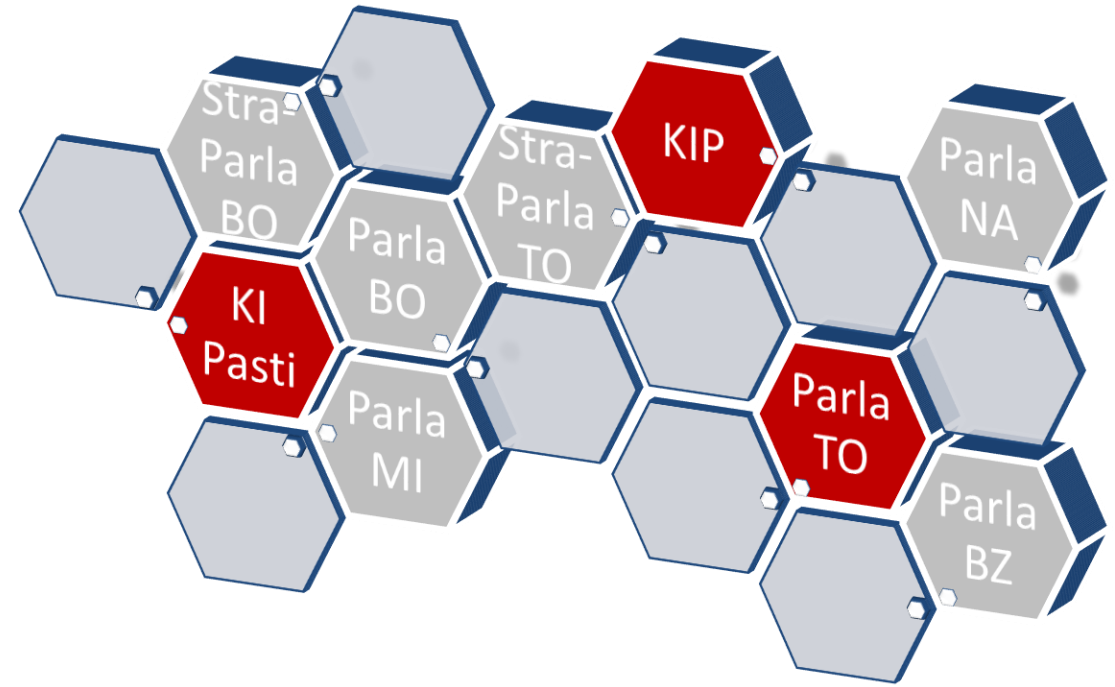
(Mauri et al. 2019)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

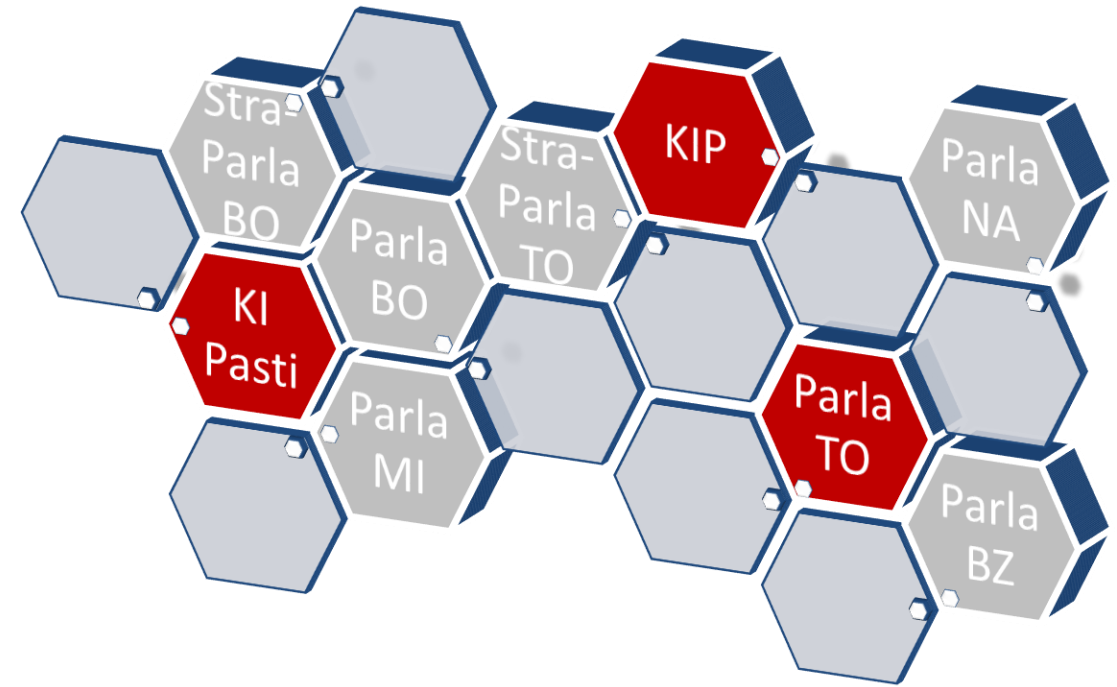
Modularità e incrementalità

- **Modularità:** divisione interna in moduli indipendenti



Modularità e incrementalità

- **Modularità:** divisione interna in moduli indipendenti
 - ✓ possibilità di consultare ogni modulo separatamente o tutti i moduli insieme;



SELECT CORPUS



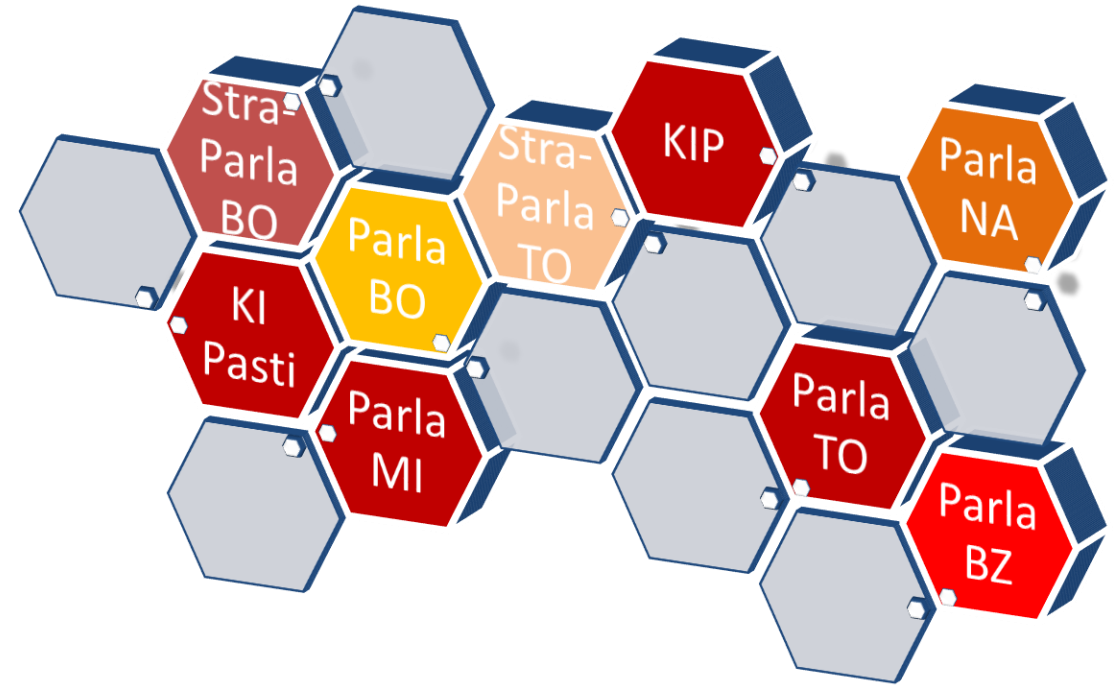
FREE CORPORA

Italian	KIP	581,784 words	OPEN
Italian	KIPARLA	1,385,219 words	OPEN
Italian	KIPasti	404,896 words	OPEN
Italian	ParlaTO	476,047 words	OPEN



Modularità e incrementalità

- **Incrementalità:** possibilità di aggiungere nuovi moduli nel tempo

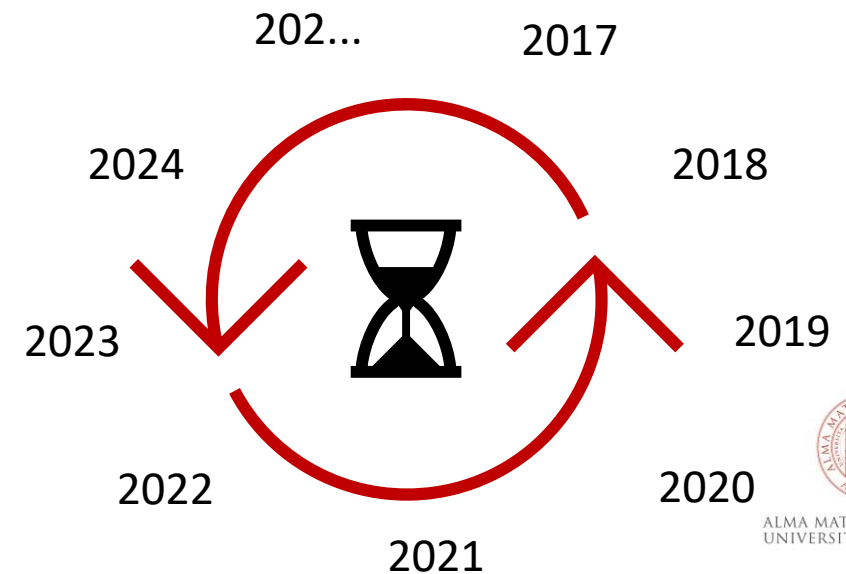
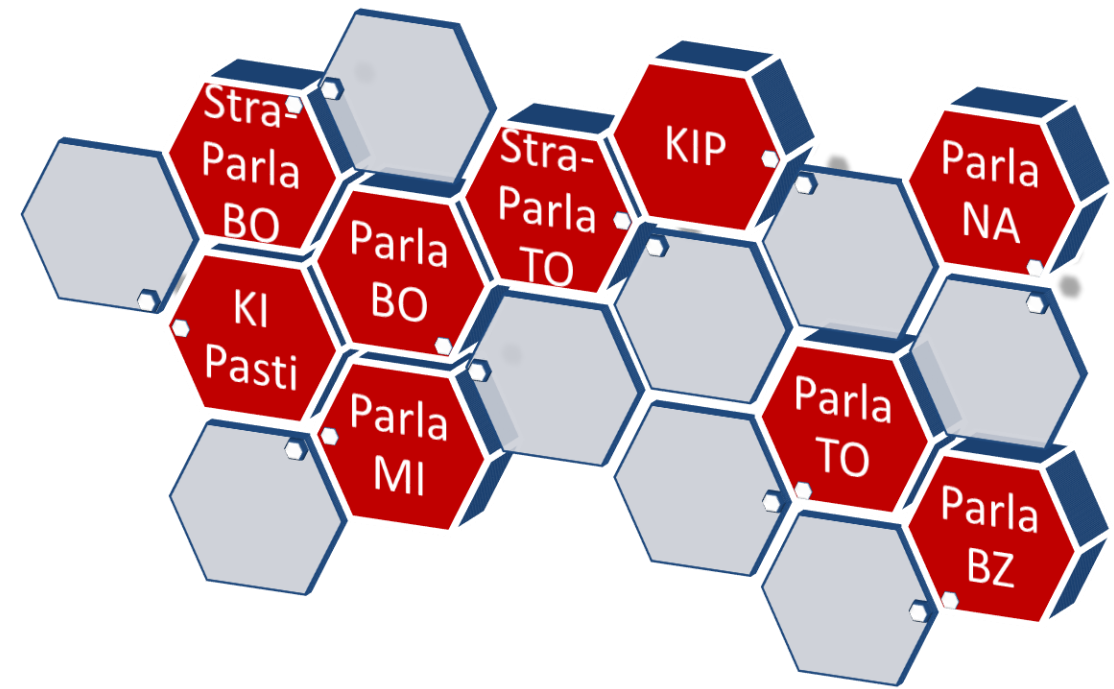


Modularità e incrementalità

- **Incrementalità:** possibilità di aggiungere nuovi moduli nel tempo

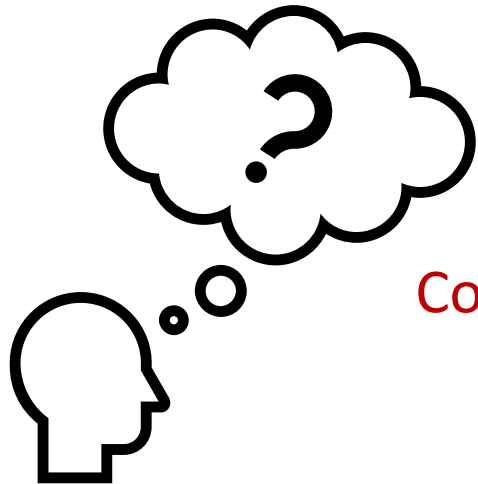
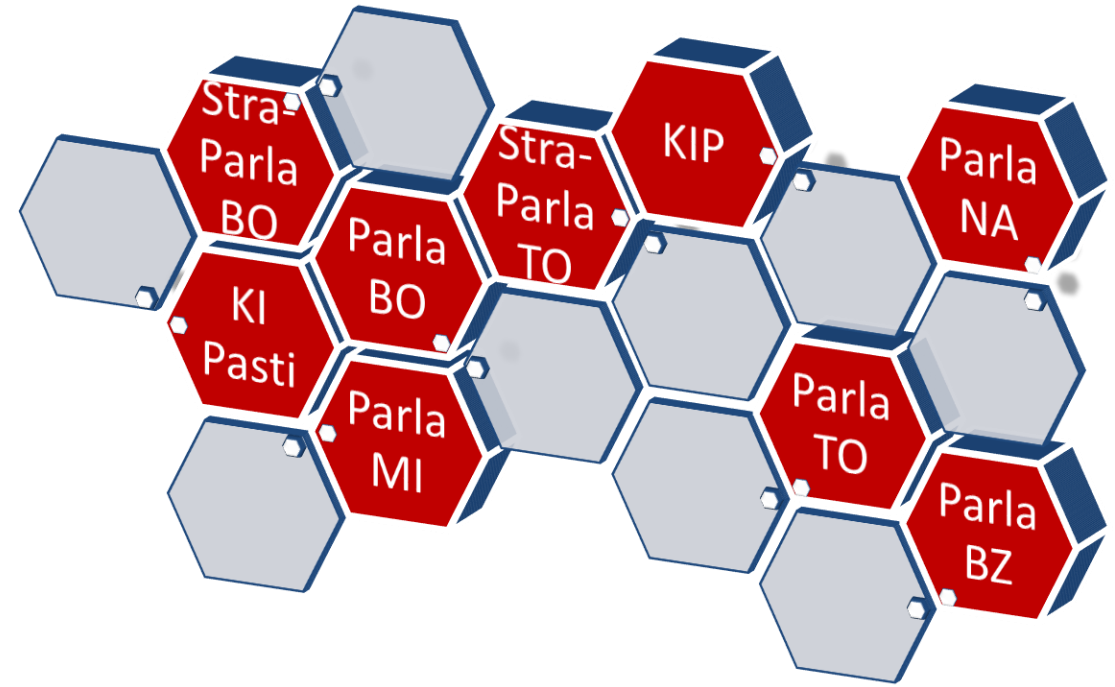


Questa natura dinamica rende il corpus KIParla adatto a documentare il linguaggio parlato nel tempo.



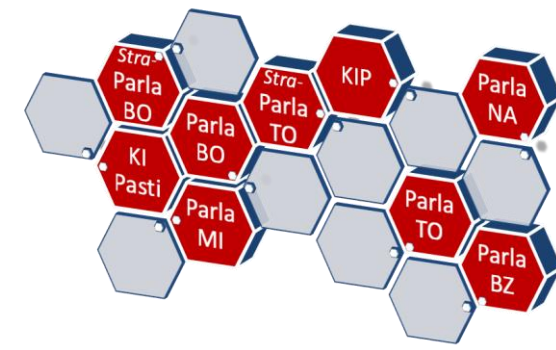
Modularità e incrementalità

- **Incrementalità:** possibilità di aggiungere nuovi moduli nel tempo



Cosa sono i **moduli** KIParla?

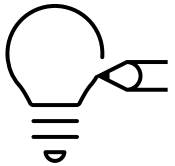
Modularità e incrementalità



I moduli sono **corpora** di piccole/medie dimensioni.

- ✓ costruiti per **scopi** diversi, spesso complementari
- ✓ dati provenienti da diverse **aree geografiche**, **tipi di interazione** e **tipologie di comunità**
- pur mantenendo **una sostanziale comparabilità di fondo nella struttura e nell'accessibilità**.

La comparabilità e allo stesso tempo la **specificità** di ogni modulo sono ciò che può rendere il corpus KIParla **rappresentativo dell'italiano parlato nel tempo**:



La visione:



più moduli vengono aggiunti
più dimensioni di variazione possono essere esplorate



Raccolta e trascrizione: *it takes a village!*



GDPR - Regolamento 2016/679

2018 - oggi: più di 100 studenti (delle università di Bologna e Torino) hanno partecipato alla costruzione del corpus KIParla. Tirocini, tesi di laurea, tesi di laurea magistrale...



Formazione e
supervisione



Aggiornamenti frequenti e
riunioni settimanali



Organizzazione e
coordinamento

- **Trascrizione** tramite ELAN, Jefferson semplificato (2004) per gli aspetti conversazionali.
- **Pseudonimizzazione:** eliminazione dei dati sensibili dalla trascrizione e dalla traccia audio.



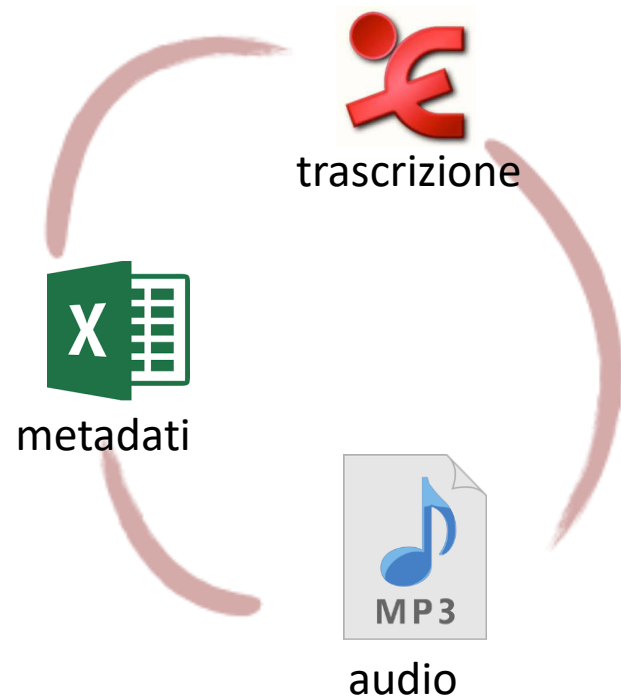
,	Intonazione crescente	<ciao>	Pronuncia (più lenta)
.	Intonazione discendente	[ciao]	Sovrapposizioni tra i diffusori
:	Suono prolungato	(ciao)	Testo di difficile comprensione
(.)	Breve pausa	xxx	Testo incomprensibile
>ciao<	Pronuncia (più veloce)	((ride))	Comportamento non verbale



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Pubblicazione dei dati e accessibilità

- XML
- Conll
- Txt (versione ortografica)
- Txt (versione colloquiale)
- Associazione occorrenza/metadati
- Allineamento con l'audio



script in python

- Standard internazionale
- Funzionalità avanzate
- Accesso libero

NoSketch Engine

Wav > mp3 con Audacity



Pubblicazione dei dati e accessibilità

✓ Vademecum

Archiviazione

Metadati Conversazioni

Metadati Parlanti

Vademecum per raccolta dati

- 1) Organizzare la raccolta: quali tipi di interazione? Quali parlanti? Dove? (tempi difficili da prevedere)
- 2) Costruzione del campione: numero di partecipanti e tipologie
- 3) Gestione dell'intervista
- 4) Indicazioni per raccolta di dati *naturally occurring*
- 5) Ore di registrazione previste
- 6) Codifica conversazioni e partecipanti KIParla
- 7) Classificazione e archiviazione metadati e audio



Pubblicazione dei dati e accessibilità

➤ Ogni codice è 'trasparente'

Vademecum

✓ **Archiviazione**

Metadati Conversazioni

Metadati Parlanti

MODULO	CONVERSAZIONI	PARTECIPANTI
KIP	BOA1001; TOA1001 (BOlogna/TORino + tipo + num. incrementale)	BO001; TO001 (BOlogna /TORino + num. incrementale)
ParlaTO	PTA001; TOD001 (ParlaTo + fascia età + num; vd.KIP)	TOR001; TOI001; TO001 (TORino + Raccoglitore + num; TORino + Intervistato + num; TORino + num, vd. KIP)
ParlaBO	PBA001 (ParlaBo + fascia età + num)	BOR001; BOI001 (BOlogna + Raccoglitore + num; BOlogna + Intervistato + num)
KIPasti	KPN001 (KiPasti + Nord/Centro/Sud + num)	PKP001 (Parlante + KiPasti + num)
Stra-ParlaBO	SBIA001; SBCA001 (Stra + Bologna + Intervista/Pasto + Arabo/Bengalese/Cinese/Ucraino + num)	PSB001 (Parlante + Stra + Bologna + num)
Stra-ParlaTO	STIA001; STCA001 (Stra + Torino + Intervista/Pasto + Arabo/Spagnolo/Cinese/Rumeno + num)	PST001 (Parlante + Stra + Torino + num)

Pubblicazione dei dati e accessibilità

Vademecum

Archiviazione

- Ogni conversazione viene inserita subito nel file Conversations, associata ai codici dei parlanti, alla sua lunghezza, a uno o più trascrittori e raccoglitori, e viene modificato lo stato della trascrizione

✓ **Metadati Conversazioni**

type	code	duration	city	community	year	participants number	p1	p2	p3	languages	Trascrittore	Raccoglitore	stato trascrizione
Intervista	SBIA001	0:19:36	Bologna	marocchina	2024	2	PSB014	PSB015		ita	ABNDHCK	****	avviata
Intervista	SBIA002	1:20:00	Bologna	marocchina	2024	3	PSB005	PSB017	PSB016	ita; marocchino	ABNDHCK	Eleonora Zucchini	avviata
Intervista	SBIA003	0:38:42	Bologna	marocchina	2024	3	PSB005	PSB023	PSB024	ita	ABNDHCK	Eleonora Zucchini	conclusa

Pubblicazione dei dati e accessibilità

Vademecum	participant code	participant occupation	participant sex	files in which participant appears	mothertongue	participant birth country (if Italy, region)	age of arrival in Italy
	PSB001	intell	F	SBIC001; SBCC001; SBCC002	cinese mandarino	cina	28
Archiviazione	PSB002	stud	F	SBIC001	cinese mandarino; huarong hua (gruppo xiang)	cina	24
	PSB003	stud	F	SBIC002; SBCC002	cinese mandarino	cina	22
Metadati Conversazioni	PSB004	stud	F	SBIC002; SBCC002	cinese mandarino	cina	27
	PSB005	intell	F	SBIC002; SBCC002	italiano	emilia-romagna	N/A
✓ Metadati Parlanti	PSB007	intell	F	SBIU001	ucraino; russo	ucraina	33

years spent in italy	participant city of residence	participant age range	participant degree	liberatoria	altre informazioni rilevanti	participant degree2
4	Bologna	31-35	post	si	dottorato in linguistica; lm in italia	phd
5	Bologna	26-30	lm	si	studentessa DAMS musica	laurea
6	Bologna	26-30	lm	si	dottoranda linguistica; lm in italia	laurea
3	Bologna	26-30	lm	si	dottoranda linguistica	laurea
N/A	Bologna	31-35	post	si		phd
2	Bologna	31-35	post	si	direttrice centro italiano in Ucraina	phd

ACCESSIBILITÀ

➤ liberamente accessibile

AUDIO

➤ Accessibile gratuitamente previa registrazione

TOD2015

☐ Anno di raccolta	2019
☐ annotation.audio_file	https://kiparla.lingue.unibo.it/kiparla/player/player.cgi?code=TOD2015&begin=509203&end=513587
☐ annotation.begin	509203
☐ annotation.end	513587
☐ annotation.files_in_which_participant_appears	TOD1004, TOD2015
☐ annotation.participant_code	TO073
☐ conversation.code	TOD2015
☐ conversation.duration	00:30:52
☐ conversation.topic	fisso
☒ doc.doc_number	TOD2015
☐ doc.full_conversation	https://kiparla.lingue.unibo.it/kiparla/conversations/TOD2015.html
☐ doc.full_jefferson	https://kiparla.lingue.unibo.it/kiparla/conversations/TOD2015_jefferson.html
☐ Età	21-25
☐ Genere	F
☐ Luogo di raccolta	TO

Audio TOD2015

Start time: 00:08:29, end time: 00:08:33, duration: 4.38 sec

II 8:26 / 30:52

🔊

⋮

TO071	((ride))
TO071	ah ma quindi tu hai sentito cose [che ha fatto jack, ah okay]
TO073	[sì sì sì (.) sì sì (.)] ciao jack se mai sentirai questo audio ((ride))
TO071	((ride))
TO073	>anche perchè camera mia e di jack era< (si:)

TO071	ti chiudevi in camera
TO073	esatto ma dall' altra parte
TO073	>sì che poi era (.) attaccata a quella di jack< quindi era come se fossi in camera insieme in realtà ((ride))
TO071	ah ma quindi tu hai sentito cose [che ha fatto jack, ah okay]
TO073	[sì sì sì (.) sì sì (.)] ciao jack se mai sentirai questo audio ((ride))
TO073	>anche perchè camera mia e di jack era< (si:)
TO073	insomma era suddivisa da un muro di cartongesso

Panoramica

1. Progettazione e implementazione del corpus

2. I moduli

3. Utilizzo del corpus: alcuni casi di studio

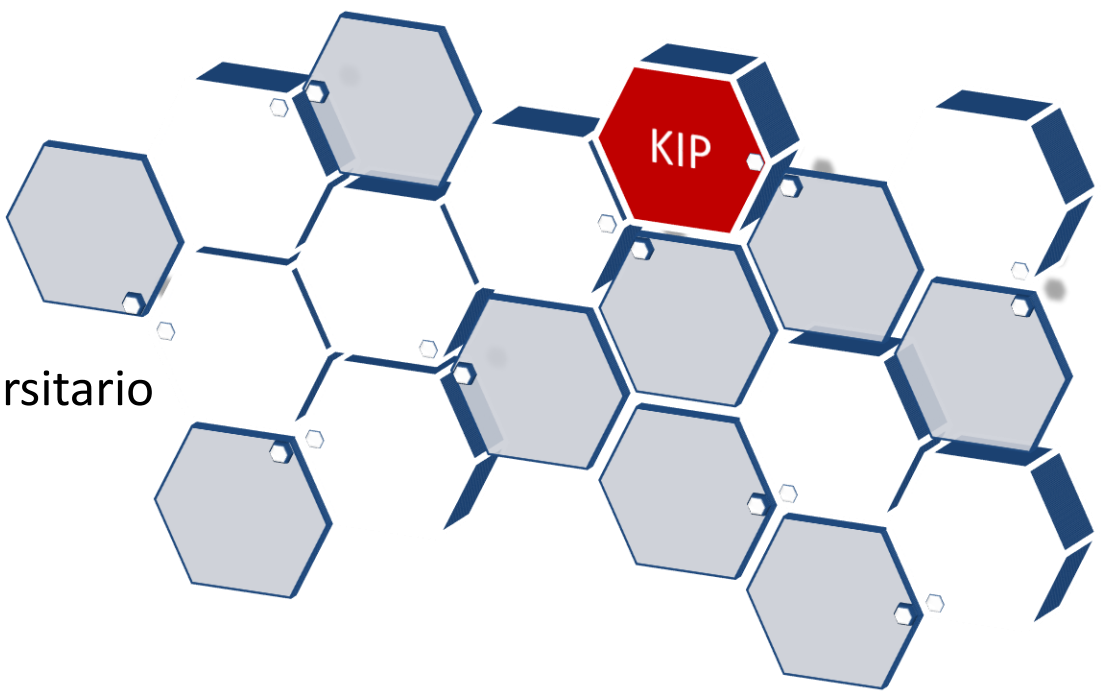


Progettazione del corpus: il modulo KIP

Modulo KIP

- 273 **relatori istruiti** registrati in un ambiente universitario
- Osservazione della variazione diafasica e regionale.

- ✓ Dati raccolti a Bologna e Torino (2016-2019)
- ✓ **Contesti comunicativi diversi**: 121 interazioni
- ✓ **TOT: circa 70 ore, 660.000 token**



	Relazione tra i partecipanti	Moderatore	Argomento
Conversazione libera	Simmetrico	Assente	libero
Intervista semi-strutturata	Simmetrico	Presente	Fisso
Ricevimento studenti	Asimmetrico	Assente	libero
Esami	Asimmetrico	Assente	libero
Lezioni	Asimmetrico	Presente	Fisso



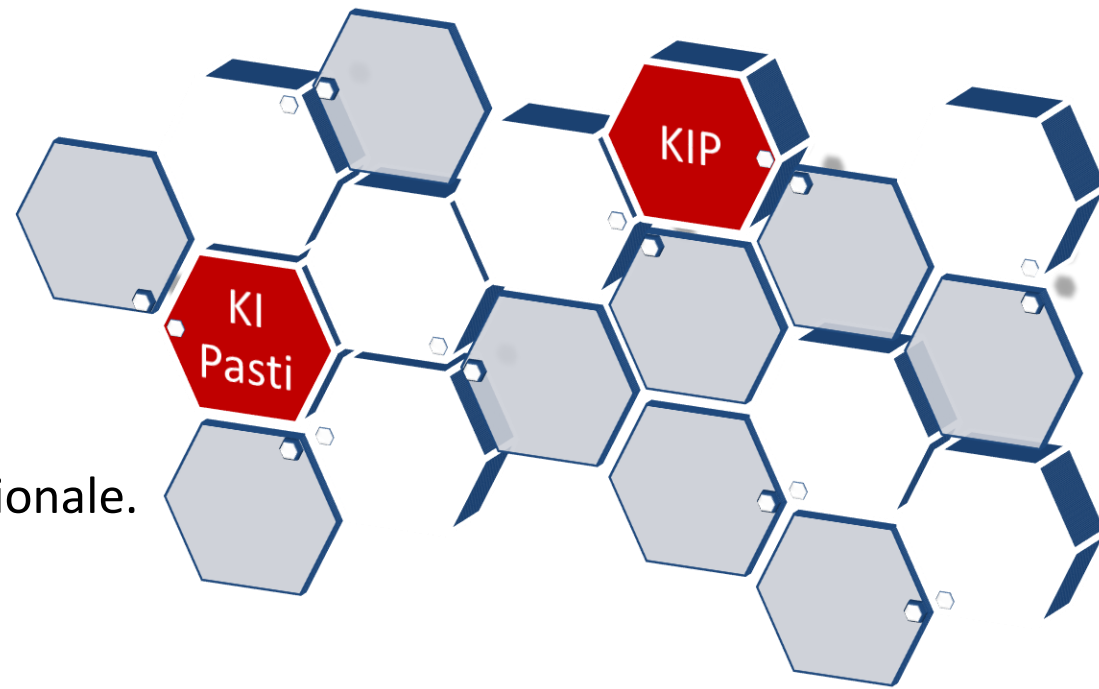
Progettazione del corpus: il modulo KIPasti

Modulo KIPasti

- 63 conversazioni a tavola in tutta Italia

→ Osservazione della variazione diafasica, diastratica e regionale.

- ✓ Dati raccolti in 13 regioni italiane (2020-2024)
- ✓ 147 Parlanti **con diversa caratterizzazione sociale** (età, titolo di studio, occupazione)
- ✓ Un contesto: **conversazioni a tavola**, quindi background condiviso, registro informale e amichevole.
- ✓ **TOT: circa 42 ore, 487.000 token**



	Raccolta dati
Italia settentrionale	46%
Italia centrale	16%
Italia meridionale	38%



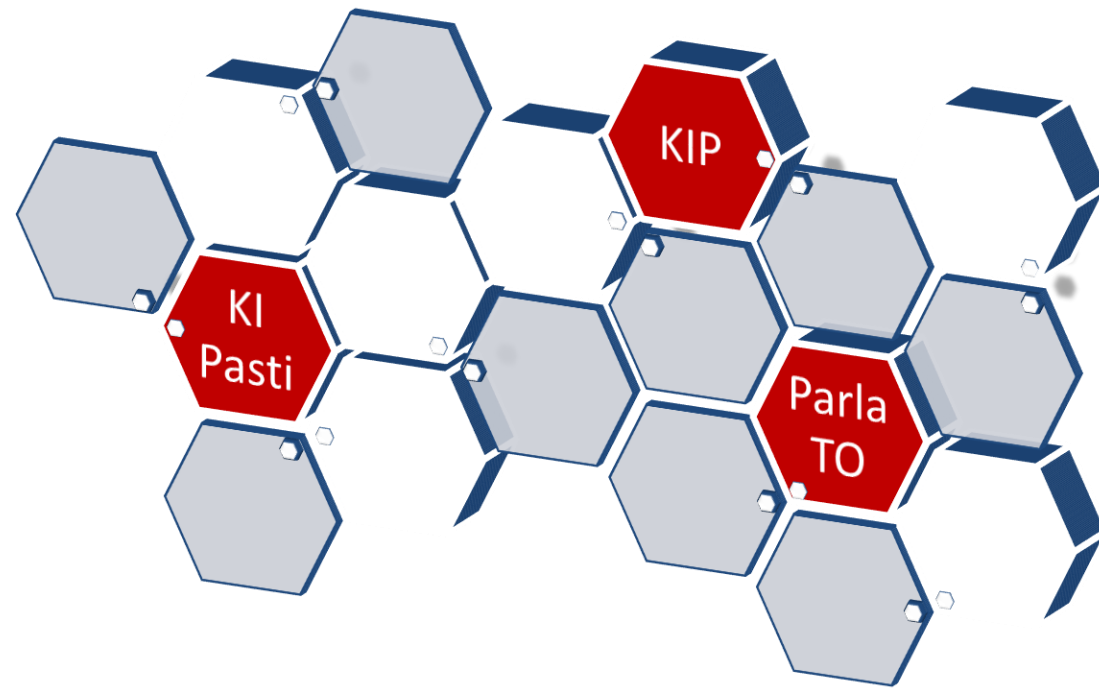
Progettazione del corpus: il modulo ParlaTO

Modulo ParlaTO

- Lingua parlata nella città metropolitana di Torino

→ Osservazione della variazione diastratica e regionale.

- ✓ Dati raccolti a Torino (2019)
- ✓ 88 Parlanti con **diversa caratterizzazione sociale** (età, grado di istruzione, occupazione)
- ✓ Un contesto: 67 interviste semi-strutturate
- ✓ **TOT: circa 50 ore, 560.000 token**



	Gruppi di età
Giovane	$18 \leq x \leq 30$ anni
Adulti	$30 < x \leq 60$ anni
Anziani	$60 < x \leq 89$ anni



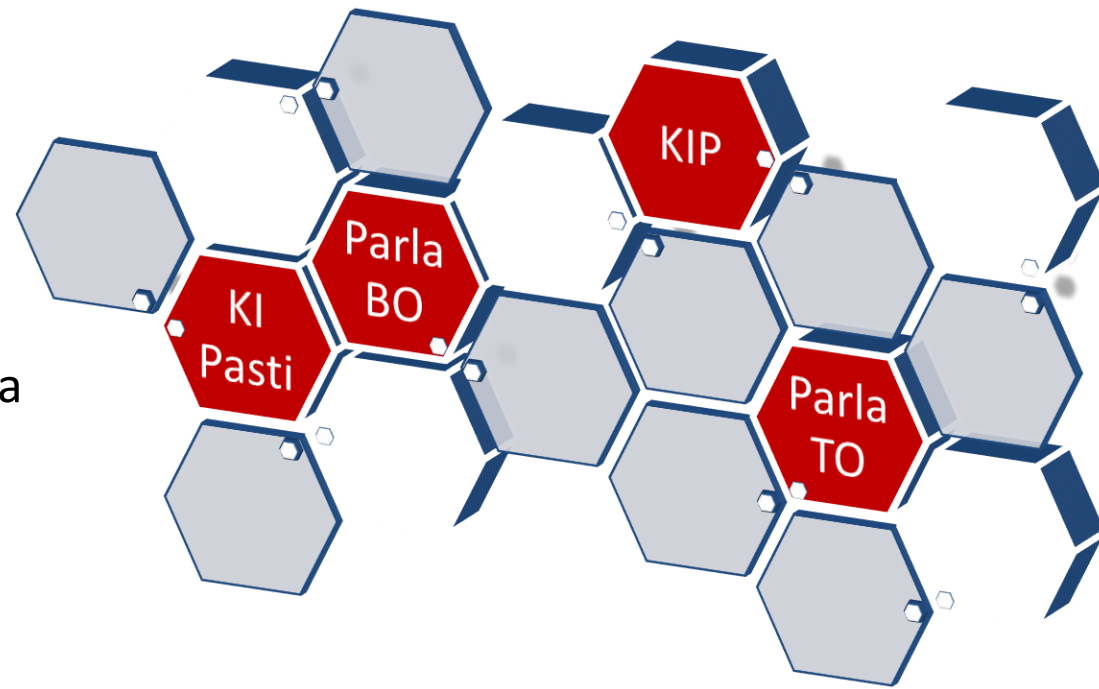
Progettazione del corpus: il modulo ParlaBO

Modulo ParlaBO

- Lingua parlata nella città metropolitana di Bologna

→ Osservazione della variazione diastratica e regionale.

- ✓ Dati raccolti a Bologna (2022-2024)
- ✓ 158 Parlanti con **diversa caratterizzazione sociale** (età, grado di istruzione, occupazione)
- ✓ Un contesto: 86 interviste semi-strutturate
- ✓ **TOT: circa 66 ore, 650.000 token**



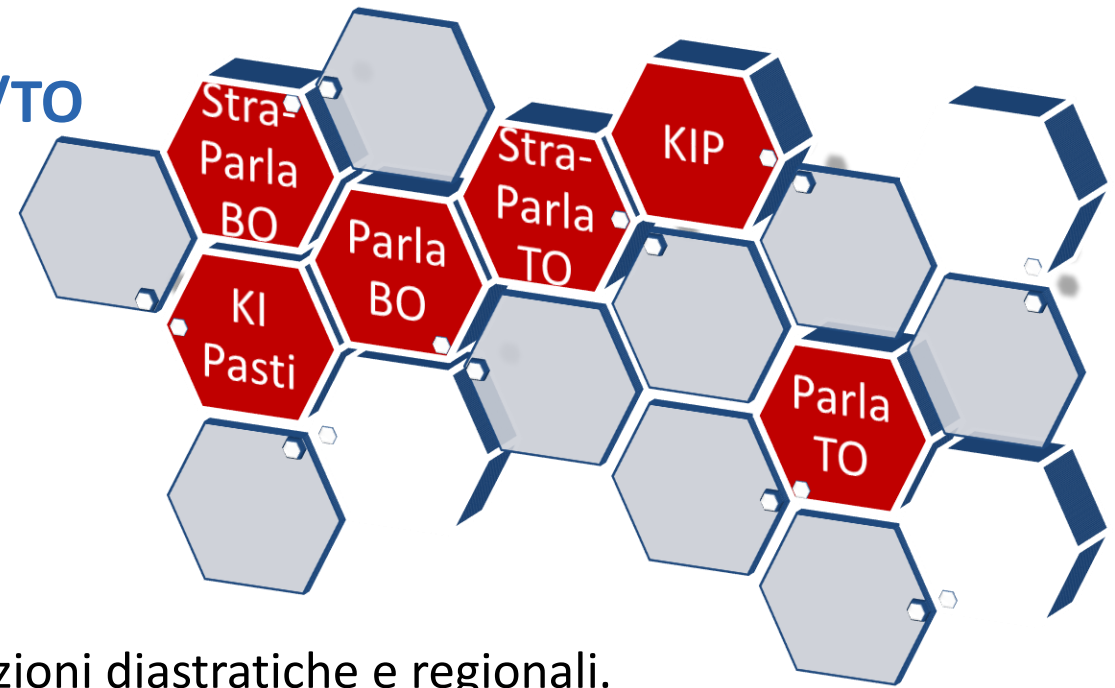
	Raccolta dati
$18 \leq x \leq 30$ anni	30%
$30 < x \leq 60$ anni	36%
$60 < x \leq 89$ anni	34%



Progettazione del corpus: Moduli Stra-Parla BO/TO

Prossimi passi: [Moduli Stra-ParlaBO e Stra-ParlaTO](#)

- L'italiano parlato nelle comunità di **parlanti di origine straniera** con repertori multilingue complessi



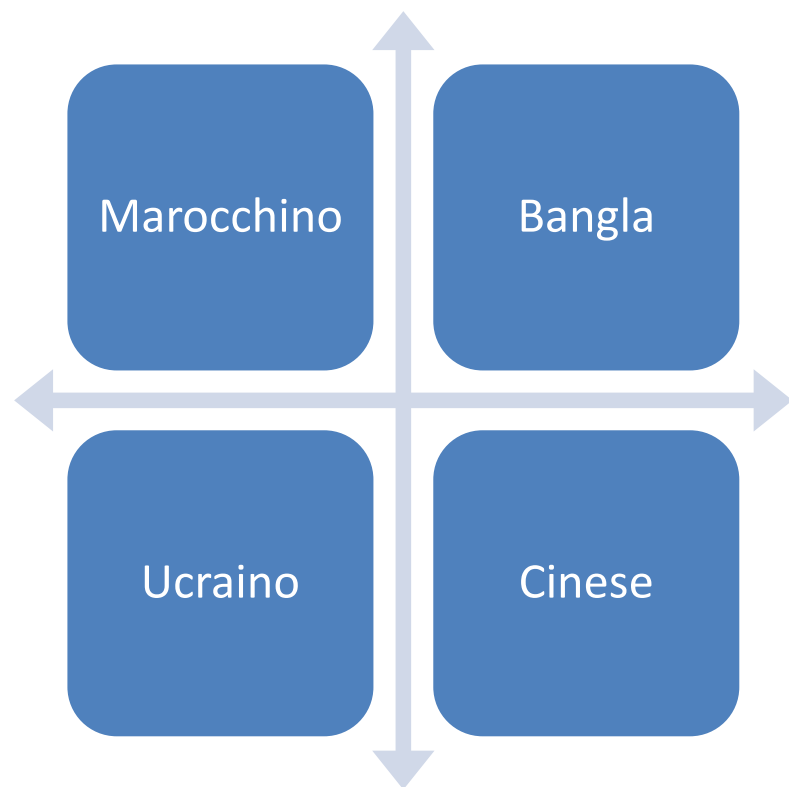
→ Osservazione delle varietà degli apprendenti, delle variazioni diastratiche e regionali.

- ✓ Dati raccolti a Bologna e Torino (2024 - in corso)
- ✓ Parlanti con **repertori multilingue, diversa caratterizzazione sociale e origine**
- ✓ Due contesti: intervista semi-strutturata, conversazione libera
- ✓ **ATTESE: circa 128 ore, 1.250.000 token**

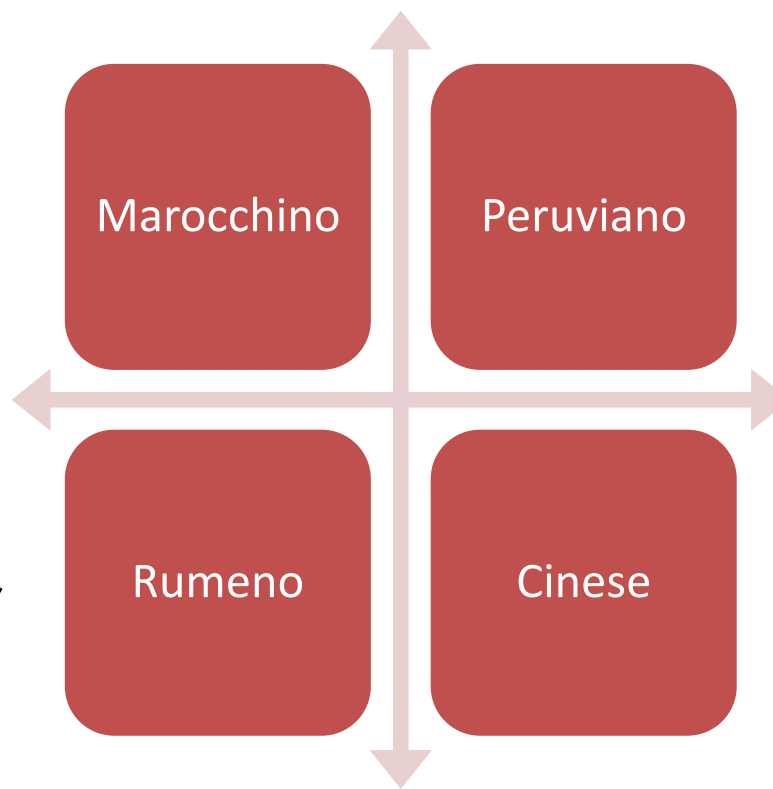
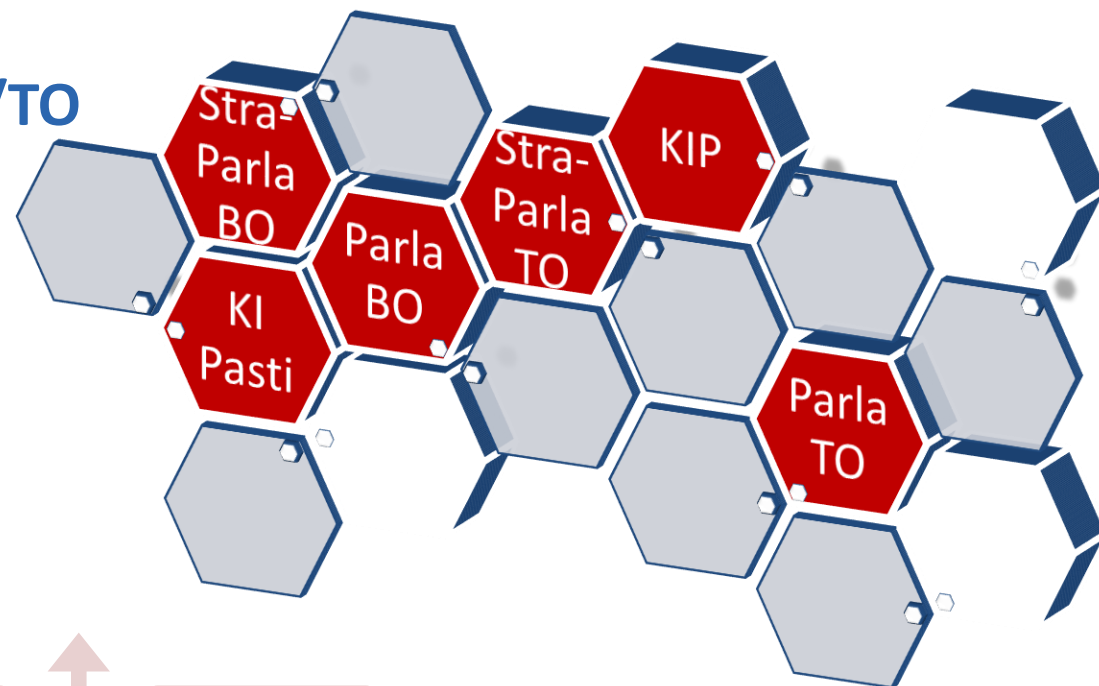
Raccolta dati in corso!



Progettazione del corpus: Moduli Stra-Parla BO/TO



Bologna



Torino

16 ore

per comunità linguistica

- 8 h Conversazione libera
- 8 h Interviste

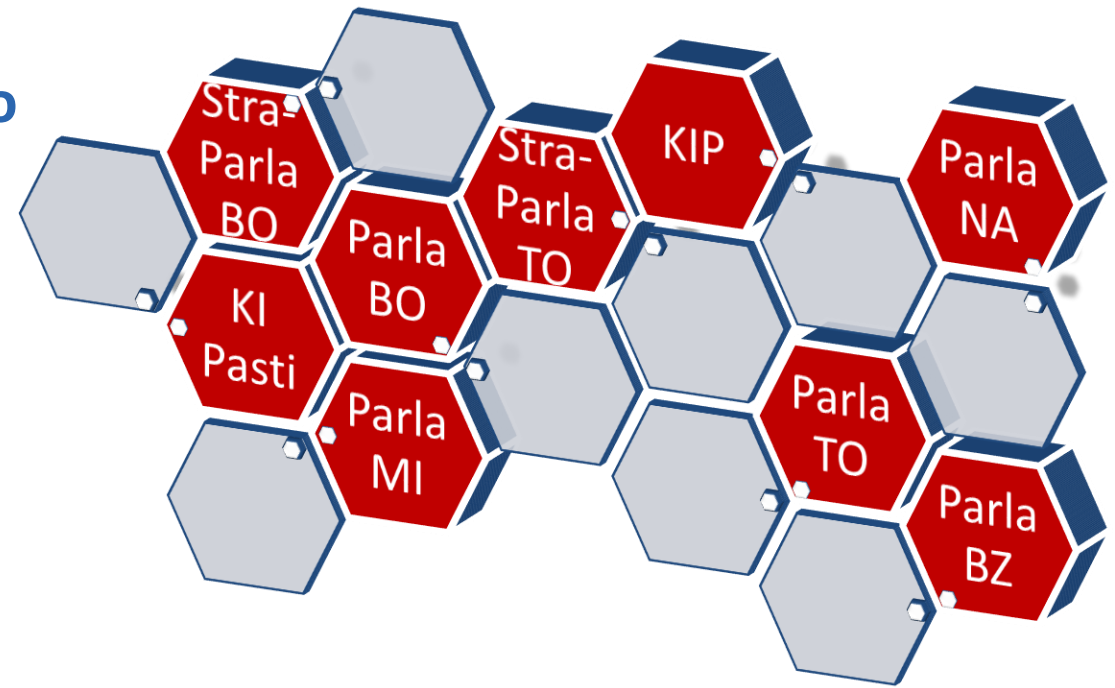


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Progettazione del corpus: collaborazioni in corso

Prossime tappe: [Moduli ParlaBZ, ParlaNA, ParlaMI](#)

- L'italiano parlato a **Bolzano** (Daniela Veronesi),
- Italiano parlato a **Napoli** (Margherita di Salvo),
- Italiano parlato a **Milano** (Federica da Milano)
- Italiano parlato a **Monaco** (Tania Paciaroni)
- Italiano parlato a **Roma**??



→ Osservazione della variazione diafasica, diastratica e regionale

Lavori in corso!



Panoramica

1. Progettazione e implementazione del corpus

2. I moduli

3. Utilizzo del corpus: prospettive ed esempi



I metadati sono filtri di ricerca

Potete costruire il vostro sottocorpus!



Text types ? ^

expand all collapse all

Tipo di interazione v

Numero di partecipanti v

Rapporto tra i partecipanti v

Presenza di moderatore v

Anno di raccolta v

Luogo di raccolta v

Partecipante v

Occupazione v

Genere v

Regione di provenienza v

Età v

Titolo di studio v

SEARCH

Text types ? ^

Tipo di interazione ^

conversazione libera

esame

intervista semistrukturata

lezione

pasto

ricevimento studenti

Età ^

16-20

21-25

26-30

31-35

36-40

41-45

46-50

51-55

56-60

SEARCH

Regione di provenienza ^

abruzzo

basilicata

calabria

campania

emilia-romagna

estero

friuli-venezia-giulia

lazio

liguria



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

I metadati sono filtri di ricerca
Potete costruire il vostro sottocorpus!

Text types

expand all collapse all

1 Tipo di interazione ▾

Numero di partecipanti ▾

Rapporto tra i partecipanti ▾

Presenza di moderatore ▾

Anno di raccolta ▾

1 Luogo di raccolta ▾

Partecipante ▾

Occupazione ▾

Genere ▾

Regione di provenienza ▾

1 Età ▾

Titolo di studio ▾

CONCORDANCE

KIPARLA

Text types 3 (3) ...

simple poi • 37

22.77 per million tokens • 0.0023%

🔍 ⬇️ 👁️ 📄 ✂️ ≡ ≡ GD EX 📄 ⋮ KWIC ⓘ

☐ Details

Left context

KWIC

Right context

1	☐	ⓘ BOD2001 ;pitante potrebbero aiutare questa cosa qua che mi dici cioè il fatto che	poi	alla fine si va all' estero ma non si conosce la gente che veramente vivi	📄
2	☐	ⓘ BOD2001 nh sì sì perché come al solito cioè se io abito in un posto non è che ho	poi	tanto interesse a conoscere chi chi viene da fuori perché cioè immagin	📄
3	☐	ⓘ BOD2006 i mi piacciono molto le luci // // eh l' importante é che le luci siano calde	poi	non sono un grande amante dell' arredamento quindi mi mi interessa pi	📄
4	☐	ⓘ BOD2006 abbia un ambiente molto confortevole // mh // quindi é quello che serve	poi	i libri me li posso procurare altrove // tranquillamente questo é solo il lue	📄
5	☐	ⓘ BOD2006 na nei periodi di lezione principalmente quindi penso // sí sí sí sí sí sí //	poi	avendo anche altri impegni la sera qui suonando eh spesso ero costret	📄
6	☐	ⓘ BOD2006 to era il fatto di avere uno spazio solo per me cioè la stanza // okay // e	poi	in realtà di poter usufruire di momenti di vita comune costantemente e i	📄
7	☐	ⓘ BOD2006 ndando per concerti // ah // preferisco spostarmi // eh in realtà in realtà	poi	dopo una giornata impegnativa // anche un bar di campagna qualche a	📄
8	☐	ⓘ BOD2006 a almeno una volta e dove vi vedete? // mah soprattutto in sala prove //	poi	in realtà avendo anche // eh avendo morose che sono amiche tra di lor	📄

(1) Variazione di registro: *per cui* lit. 'per cui'

CONCORDANCE KIPARLA

simple *per cui* • 438
269.58 per million tokens • 0.027%

Left context KWIC Right context

1	☐	☐	KPN015	icare bene un vino // lo sono anche fra i sommelier // <i>per cui</i> il sommelier che nasce come categoria comunicativa
2	☐	☐	KPN015	ibera // si sentiva libera di dire quello che le veniva // <i>per cui</i> era capace di dirti quel vino qua sa di calzini // // cik
3	☐	☐	BOA3004	attivato le reminescenze del capodanno di grignani // <i>per cui</i> io mi sono messa a fare la cazzona imitando // però c
4	☐	☐	BOA3004	ito che non superano un certo valore economico indi <i>per cui</i> so' veramente piccolezze // cioè a me già non interes
5	☐	☐	BOA3004	laurea e dall' idoneità d' inglese // idoneità d' inglese <i>per cui</i> ho studiato zero // ho provato a far finta c' è io c' ho p
6	☐	☐	BOA3004	l' altra // ne ho fatto un altro che si chiama un amore <i>per cui</i> lottare // oh mio dio // lei figlia di celerini // oh yes // lu



(1) Variazione di registro: *per cui* lit. 'per cui'

✓ *Per cui* ha subito un processo di grammaticalizzazione:

- **dalla** costruzione obliqua relativa "per cui" (REL)

(1) KIP, TOD1014

TO068: quindi vi chiedo // eh vabbè mh innanzitutto

il **motivo per cui** avete deciso di frequentare questo corso

CONCORDANCE		KIPARLA	
simple <i>per cui</i> • 438 269.58 per million tokens • 0.027%			
☐ Details	Left context	KWIC	Right context
1	☐ ⓘ KPN015	icare bene un vino // lo sono anche fra i sommelier //	<i>per cui</i> il sommelier che nasce come categoria comunicativa
2	☐ ⓘ KPN015	ibera // si sentiva libera di dire quello che le veniva //	<i>per cui</i> era capace di dirti quel vino qua sa di calzini // //
3	☐ ⓘ BOA3004	attivato le reminescenze del capodanno di grignani //	<i>per cui</i> io mi sono messa a fare la cazzona imitando // però c
4	☐ ⓘ BOA3004	ito che non superano un certo valore economico indi	<i>per cui</i> so' veramente piccolezze // cioè a me già non interes
5	☐ ⓘ BOA3004	laurea e dall' idoneità d' inglese // idoneità d' inglese	<i>per cui</i> ho studiato zero // ho provato a far finta c' è io c' ho p
6	☐ ⓘ BOA3004	l' altra // ne ho fatto un altro che si chiama un amore	<i>per cui</i> lottare // oh mio dio // lei figlia di celerini // oh yes // lu

Metadati TO068

Origine: Sicilia (Italia)

Età: 36-40 anni

Titolo: Dottorato di ricerca

Occupazione: Professionisti



(1) Variazione di registro: *per cui* lit. 'per cui'

✓ *Per cui* ha subito un processo di grammaticalizzazione:

- **dalla** costruzione obliqua relativa "per cui" (REL)

(1) KIP, TOD1014

TO068: quindi vi chiedo // eh vabbè mh innanzitutto

il **motivo per cui** avete deciso di frequentare questo corso

- **al** connettivo conclusivo "quindi" (CONN)

(2) ParlaTO, PTD012

TOR001: certo

TOI068: **per cui** *diciamo* che sì

boh non so dimmi che cosa vuoi che ti racconti

CONCORDANCE		KIPARLA	
simple per cui • 438		269.58 per million tokens • 0.027%	
☐	Details	Left context	KWIC
1	☐ ⓘ KPN015	icare bene un vino // lo sono anche fra i sommelier //	per cui il sommelier che nasce come categoria comunicativa
2	☐ ⓘ KPN015	ibera // si sentiva libera di dire quello che le veniva //	per cui era capace di dirti quel vino qua sa di calzini // //
3	☐ ⓘ BOA3004	attivato le reminescenze del capodanno di grignani //	per cui io mi sono messa a fare la cazzona imitando // però c
4	☐ ⓘ BOA3004	ito che non superano un certo valore economico indi	per cui so' veramente piccolezze // cioè a me già non interes
5	☐ ⓘ BOA3004	laurea e dall' idoneità d' inglese // idoneità d' inglese	per cui ho studiato zero // ho provato a far finta c' è io c' ho p
6	☐ ⓘ BOA3004	l' altra // ne ho fatto un altro che si chiama un amore	per cui lottare // oh mio dio // lei figlia di celerini // oh yes // lu

Metadati TO068

Origine: Sicilia (Italia)

Età: 36-40 anni

titolo: Dottorato di ricerca

Occupazione: Professionisti

Metadati TOI068

Origine: Lazio (Italia)

Età: 46-50 anni

titolo: Laurea

Occupazione: Professionisti

(1) Variazione del registro: *per cui*

L'uso non relativo di *per cui* è oggi **accettato**, ma principalmente associata alle interazioni **informali e parlate**.

1879

Tommaseo:
Non deve essere utilizzato

Dizionario della lingua italiana:
Per cui a modo d'avverbio, in senso di *Per la qual cosa*, *Dunque*, non è da dire.

1969

Battaglia:
questo uso non è corretto

- *Per cui*: con valore di neutro, nel significato di *per la qual cosa*, *per ciò*, *per questo* (ma non è uso corretto).

Guarini, 80: Questa rimembranza / ... / è quasi un agitar fiaccola al vento, / *per cui*, quanto l'incendio / sempre s'avanza, tanto / a l'agitata fiamma ella si strugge.

2008

Treccani:
soprattutto nel linguaggio parlato

"spesso *per cui* è usato, spec. nel linguaggio parlato, col valore della congiunzione conclusiva e *perciò*".

2024

Sabatini & Coletti:
non è adatto al registro tecnico-formale scritto

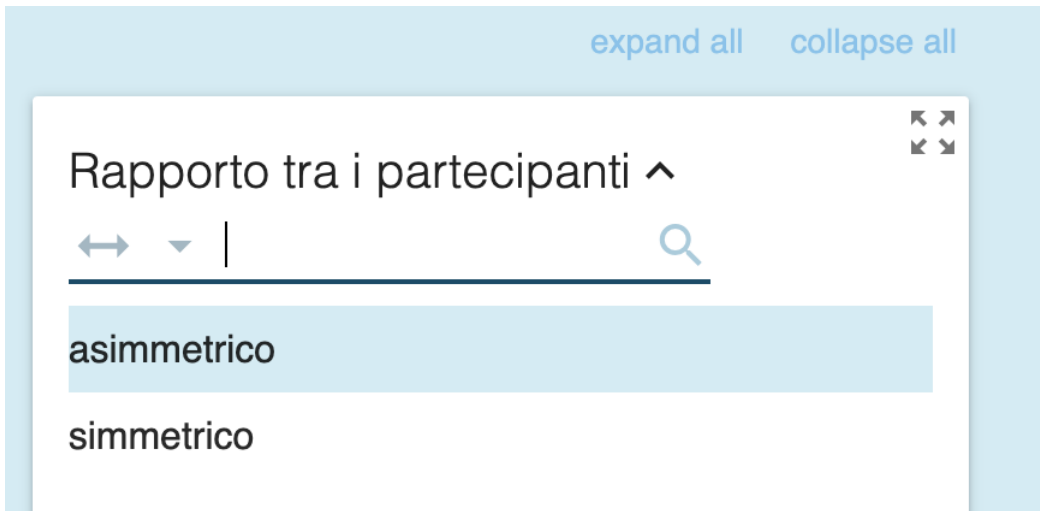
"*Sicché, perciò*. Frequente nel parlato e nelle scritture di livello stilistico medio, non si aggiunge alle scritture di carattere tecnico-formale".

➤ Variazione sociolinguistica?



(1) Variazione del registro: *per cui*

- KIP: **conversazioni spontanee**, **interviste semi-strutturate**, **ore di ufficio**, **lezioni**, **esami**;
- ParlaTO: **interviste semi-strutturate**;
- KIPasti: **conversazioni a tavola**;



✓ **Registro formale**
TOT: 361.702 tokens

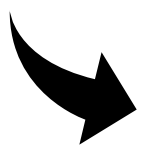
✓ **Registro informale**
TOT: 1.339.021 tokens

(1) Variazione del registro: *per cui*

	REL	CONN	TOT.
Registro informale	30 (35,3%)	55 (64,7%)	85
Registro formale	70 (20,2%)	277 (79,8%)	347

Il valore statistico del test esatto di Fisher è 0,0041. Il risultato è **significativo** a $p < .01$.

- ✓ Gli usi connettivi di *per cui* sono **significativamente più frequenti** nelle **interazioni formali!**



Ballarè & Mauri 2024: la
grammaticalizzazione del *per cui*
come connettivo può essere
analizzata come un cambiamento
multifase

(Ballarè & Mauri 2024)

Iniziato come un **cambiamento dal basso**
(stigmatizzato in passato)

Continua come **cambiamento dall'alto** (nei registri formali)



(2) Congiuntivo vs. indicativo nei condizionali controfattuali

CONGIUNTIVO

PROTASI

CONDIZIONALE

APODOSI

(3) ParlaTO, PTA18

TOI031: *ovviamente se poi **avesse partorito**
ci saremmo dovuti fermare in ambulanza*

INDICATIVO - imperfetto

PROTASI

APODOSI

(4) ParlaTO, PTD006

TOI055: *se non **trovavo** niente **m' attaccavo** al tram*

ITALIANO STANDARD
(Berruto 1983)

Italiano non standard

Mauri et al. 2023

Condizionali controfattuali:

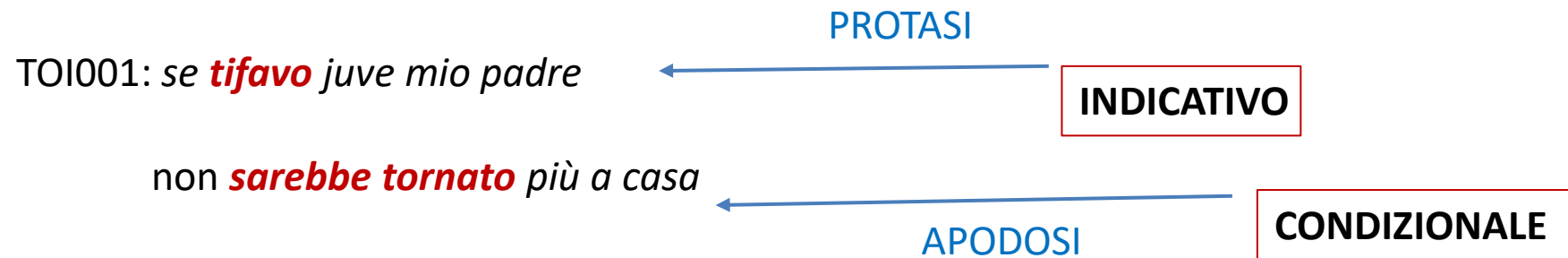
Variazione linguistica in italiano e non solo

(2) Congiuntivo vs. indicativo nei condizionali controfattuali

➤ VARIE costruzioni controfattuali non standard...

Casi misti!

(5) ParlaTO, PTA001



(2) Congiuntivo vs. indicativo nei condizionali controfattuali

➤ VARIE costruzioni controfattuali non standard...



Qual è la
distribuzione
effettiva di
**controfattuali non
standard?**

(5) ParlaTO, PTA001

TOI001: *se **tifavo** juve mio padre*

PROTASI

INDICATIVO

*non **sarebbe tornato** più a casa*

APODOSI

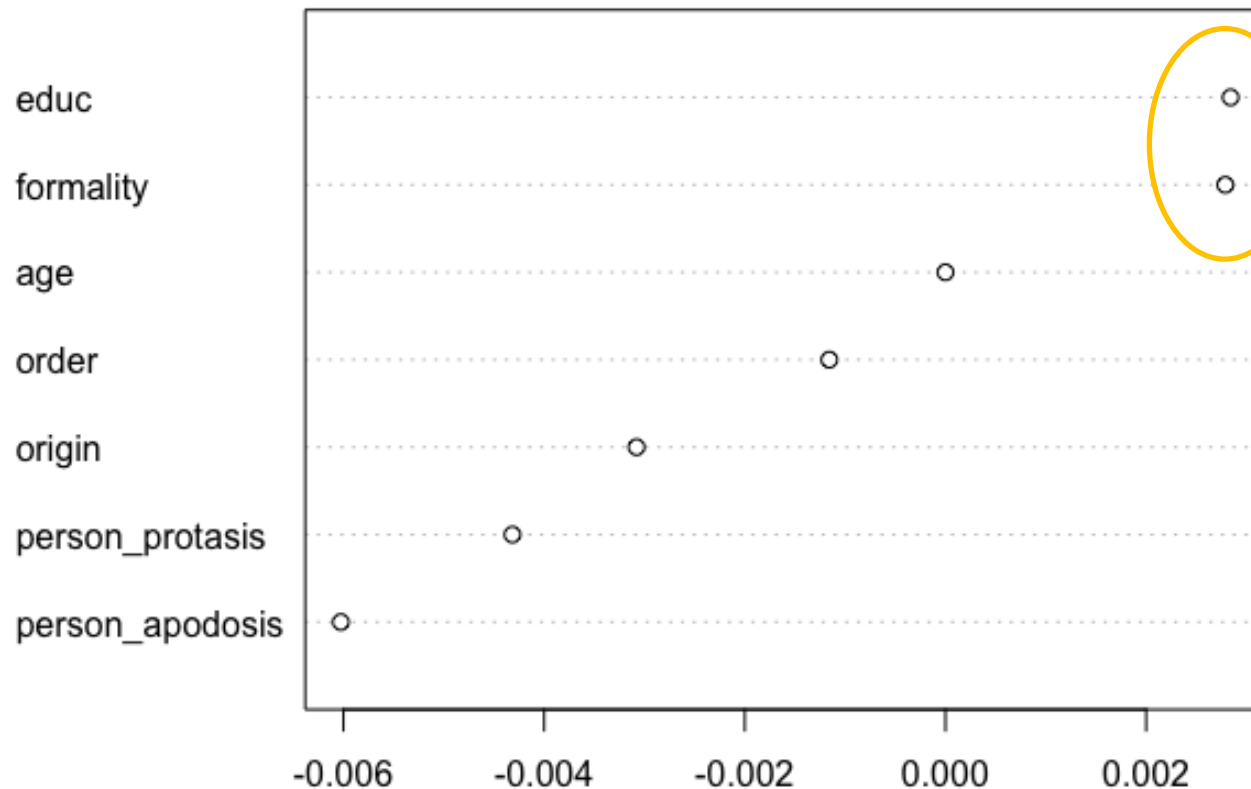
CONDIZIONALE

Mauri et al. 2023
*Condizionali controfattuali:
Variazione linguistica in italiano e non solo*

(2) Congiuntivo vs. indicativo nei condizionali controfattuali

- Tecnica della random forest

Uso dell'indicativo nei controfattuali



Goodness of fit: C= 0.8014385.

Livello di istruzione

Metadati dei parlanti

Variazione del registro

Metadati della situazione

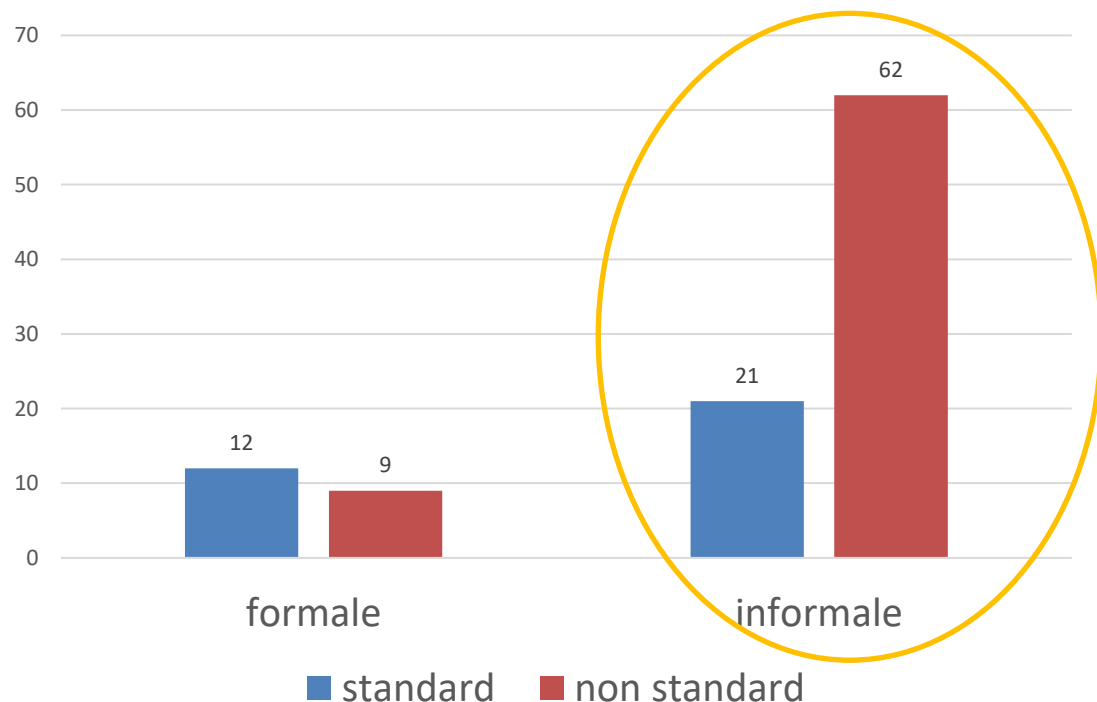
Mauri et al. 2023

Condizionali controfattuali:

Variazione linguistica in italiano e non solo

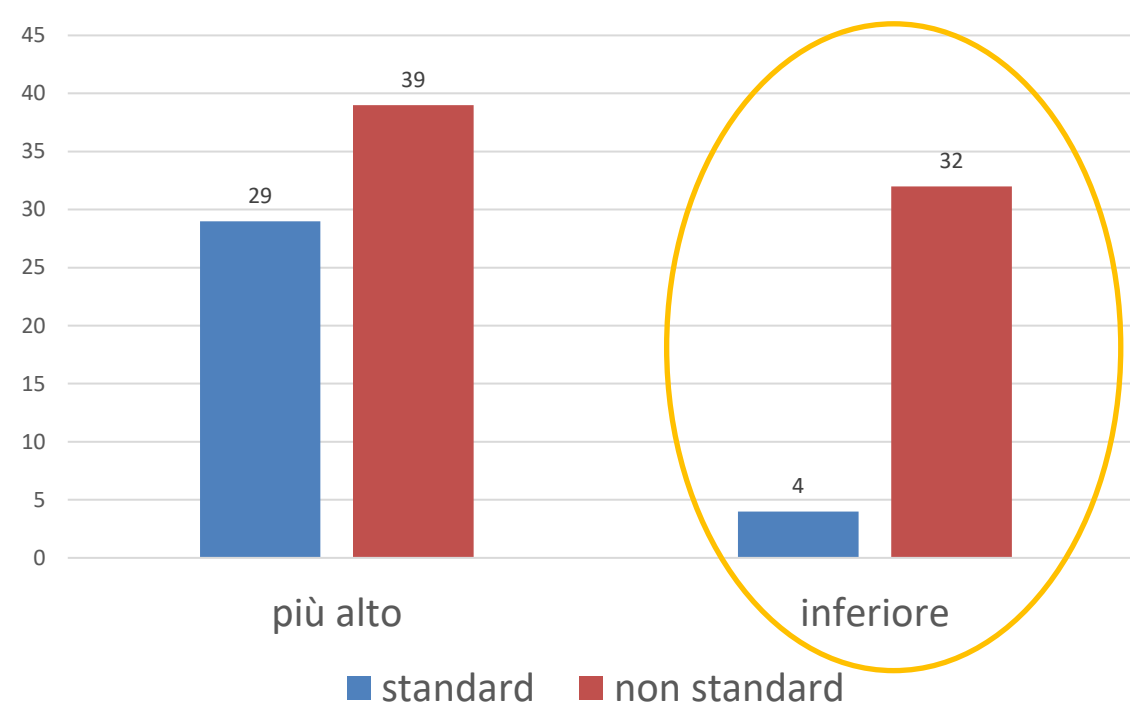
(2) Congiuntivo vs. indicativo nei condizionali controfattuali

Strategia non standard e
registro formale e informale



Secondo il test esatto di Fisher, il risultato è significativo a $p < .01^{**}$, con $p = < 0.0082$, $N=104$.

Strategia non standard e
titolo di studio dei parlanti



In base al test esatto di Fisher, il risultato è significativo a $p < .001^{***}$, con $p = < 0,0009$, $N=104$.



(3) Costruzioni emergenti: [come N]

Metadati BOI114

Origine: Emilia-Romagna (Italia)

Età: 21-25 anni

Occupazione: Studente universitario

(6) KIP, BOD2021

BO014: di dublino mi e' piaciuto vabbe' il centro
eh [[mi ricordava un po' di bologna]_{Pred} come

TO999: mhmh

BO014: **come** architettura]

➤ Qual è la funzione di **come**?

✓ **RIDUZIONE DEL DOMINIO**: in (6) l'affermazione *Dublino [...] mi ha ricordato un po' Bologna* dovrebbe essere riferita al solo **ambito** dell'*architettura*.

➤ **Marcatore di TOPIC** dislocato a destra!

Mauri (2022)



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

(3) Costruzioni emergenti: [come N]

Metadati BO156, BO158

Origine: Emilia-Romagna (Italia)

Età: 21-25 anni

Occupazione: Studente universitario

(7) KIP, BOA3021

BO158: no fidati se lo conoscessi sarebbe divertente
perché [[lui è proprio divertente]_{Pred} come persona]

?!?

(8) KIP, BOA3021

BO156: non tanto per pe per il fatto che son tossici ma per
il fatto che stanno insieme [...]
che [[non va bene]_{Pred} come cosa]

?!?



- **Semantica vuota!** *Persona* e *cosa* sono **iperonimi** generici del referente a cui si riferisce il predicato:
 - è solo divertente come persona = è solo divertente
 - Non va bene come cosa = non va bene

➤ **NESSUNO** restringe il campo d'azione o indica il topic



(3) Costruzioni emergenti: [come N]

(7) KIP, BOA3021

BO158: no fidati se lo conoscessi sarebbe divertente
perché [[lui è proprio divertente]_{Pred} come persona]
?!?

(8) KIP, BOA3021

BO156: non tanto per pe per il fatto che son tossici ma per
il fatto che stanno insieme [...]
che [[non va bene]_{Pred} come cosa]
?!?

Metadati BO156, BO158

Origine: Emilia-Romagna (Italia)

Età: 21-25 anni

Occupazione: Studente universitario

➤ **Generica ripresa anaforica** volta a sottolineare la **rilevanza** di ciò che si sta dicendo:

✓ **[Pred [come *N_hypernym*]]** sta sviluppando la funzione di **marcatore di rilevanza** con **valore di rinforzo**



(4) Indagare il posizionamento dei parlanti e la (co)costruzione delle categorie

(6) ParlaBO, PBC007

Arrivo delle truppe alleate in Romagna, durante la seconda guerra mondiale

BOI019: eh sai l'inghi- l'inghilterra era una potenza anche::

nel nel in i:ndia dalle parti infatti lì [c'erano degli india:ni c'era

m- marocchini c'era tutta la gente così]

- 
- ✓ Elenco di esempi
 - ✓ Deixis
 - **Stereotipi!**



Categoria ad hoc
[IMMIGRATI]

Metadati BOI019

Regione di nascita: Emilia-Romagna
Età: 91-95
Istruzione: Diploma
Occupazione: Pensionato



(5) Indagini sulla variazione geografica e sui dialetti



Text types (2) ? ^ expand all collapse all

Numero di partecipanti v	Anno di raccolta v	Provincia di raccolta v
Lingue conversazione ^ italiano-dialetto x 	Partecipante v	Regione di raccolta ^ umbria x 
italiano		abruzzo basilicata calabria campania emilia-romagna lazio

GO

(7) KIPasti, KPC009

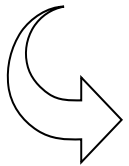
PKP089: [°ma° diec']euro se vai a mangia' da qualche parte (.) ma **magni de brutto** eh: , (.) cioè: ((ride))

PKP088: ((cavalcare))

PKP089: **#[eh so' ita] so' ita a paghe':** (.) >gli ho fatto< no: da:i pago io: perchè se no pagate sempre voi (.) m'avete fatto veni' a ce:na (.) a pranzo e:h

PKP088: [mh.]

PKP089: **#so' ita a paghe' citti:.** (.) ((respiro)) [QUANTO]?



: code-switching

✓ Uso dell'italiano regionale e del dialetto

Metadati PKP089

Origine: Umbria (Italia)

Età: 21-25 anni

titolo: Formazione professionale

Occupazione: artigiano

(6) Indagini sul gergo

(8) KIPasti, KPN022

PKP070: i pulotti possono arrivare lo ste[sso]

PKP076: [mh]mh quello sì

PKP075 i?

PKP070 pulotti

Occupazione ▾

Età ^

16-20 × 21-25 ×

26-30

31-35

36-40

41-45

46-50

51-55

56-60

61-65

Metadati PKP070

Origine: Veneto (Italia)

Età: 21-25 anni

Titolo di studio: scuola superiore

Occupazione: Studente universitario

Metada PKP076

Origine: Veneto (Italia)

Età: 56-60 anni

Laurea: Scuola media

Occupazione: Lavoratori artigianali

Metadati PKP076

Origine: Veneto (Italia)

Età: 16-20 anni

Laurea: Scuola media

Occupazione: Studente



(7) StraParlaBO: caratteristiche delle varietà degli apprendenti

➤ **Comparabilità con altre varietà di italiano!**

Ad esempio, informale o non standard.

(9) SBIB002

PSB026: ehm **un** cliente **è venuto** qua mio **\$** negozi **lasciato** una bici



\$: parola inesistente

(10) SBIB009

PSB058: ah poi diverso lavoro cercare **lavoro** difficile

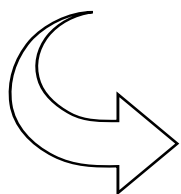
PSB005: qui (.) o in Bangladesh?

PSB058: no: bangladesh tanti **lavori** // **tanti lavori** c'è

PSB005: qui è più difficile?

PSB058: sì sì cercare difficile **lavoro**

ma il Bangladesh lavora (.) soldi di meno



- ✓ Organizzazione della sintassi per argomenti
- ✓ Nessuna distinzione netta in PoS
- ✓ Costruzioni esistenziali
- ✓ Morfologia verbale e uso
- ✓ Accordo e marcatura di genere

Metadati PSB026; PSB058

Origine: Bangladesh

Lingua madre: Bangla

Età: 36-40; 16-20

Anni trascorsi in Italia: 14; 3

Laurea: Università; Scuola media

Occupazione: Addetto alle vendite;
studente di formazione professionale

(8) StraParlaBO: commutazione di codice I



(11) SBIC001

PSB002: ci- cina ci sono otto ((ride)) sistemi

PSB001: ah sì

PSB002: di:: >come si dice< di modo [modi] [e:] cucinare

PSB001: [mhmh] [sì]

PSB002: #_suan leba zhong shuofa

PSB001: #_ah, bazhong wo yiwei ni shuo caixi

PSB002: #_dui caixi wo xiang shuo

PSB001: s::ì

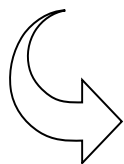
PSB002 #_dan caixi wo bu zhidao zenme qu fanyi ba
zhong shuofa yinggai ye buneng zheme

PSB001: otto filoni di

PSB005: mhmh

PSB001: cioè di cucina

PSB005: ok



- Commutazione di codice tra partecipanti e discorso
- Focus sulla forma
- Mediazione linguistica

Metadati PSB001; PSB002

Origine: Cina

Lingua madre: Cinese mandarino; Xiang

Età: 26-30; 31-35

Anni trascorsi in Italia: 4; 5

titolo: Dottorato di ricerca; laurea triennale

Occupazione: Professionisti; Studenti

(8) StraParlaBO: commutazione di codice II

(12) SBIA013

PSB005: qual è la città più bella del marocco?

PSB069: no no agadir

PSB005: ok

PSB069: la città grande

PSB067: #_ zwina agadir? // no per me no

PSB069: perché mh

PSB068: #_ ma 'mri mshit liha

PSB067: #_ min casablanca

PSB069: non è non è (.) no // non è abita (.) solo

PSB067: #_ wash kat 'ajbek agadir? zwina? hsen men casa?

PSB069: #_ batatan

PSB068: #_ ghadi nt'assab ((cavalcata))

PSB066: ((cavalcare))

PSB067: no no la più bella è casablanca

PSB005: ok



Metadati PSB067; PS069

Origine: Marocco

Lingua madre: Arabo marocchino;
francese

Età: 41-45; 46-50

Anni trascorsi in Italia: 21; 17

titolo: diploma

Occupazione: disoccupato



- Commutazione di codice
legata al partecipante



simple **grazie** • 772

475.15 per million tokens • 0.048%

☐ Details

Left context

KWIC

Right context

1	☐	ⓘ KPN015	ion è uguale dappertutto // mh xx // però eh però lì partivan dal secco //	grazie	// beh allora forse cambia
2	☐	ⓘ KPN015	biglietto // eh infatti volevo scriverti un altro biglietto con scritto non so	grazie	ben fatto // ti tengo d'occh
3	☐	ⓘ KPN015	io' di agrumato nel profumo ecco // cosa vuoi? ah // un'altra clementina	grazie	// c'ha del profumo c'ha d
4	☐	ⓘ BOA3004	nticinque // posso invitarvi a questa confusione a' // lascio il volantino //	grazie	// che cos' è maria? // è ic
5	☐	ⓘ BOA3004	ii // che però non è // ma lui ha parlato al // xxx beautiful girl // ciao // no	grazie	niente // don't worry be ha
6	☐	ⓘ BOA3004	' xxx beautiful girl // ciao // no grazie niente // don't worry be happy // no	grazie	non abbiamo niente // no
7	☐	ⓘ BOA3004	sso a posto tutto l' armadio // ciao xx // ciao // a posto? // niente guarda	<u>grazie</u>	// sì non abbiamo niente /
8	☐	ⓘ BOA3004	zie // sì non abbiamo niente // xx xx orologi // eh lo so ma abbiamo tutto	grazie	// mi spiace // xx xx c' ho s
9	☐	ⓘ BOA3004	enza nostra // ma per cosa? // è un proge~ è una ricerca di linguistica //	grazie	per avere accettato di par
10	☐	ⓘ BOA3004	untamento // ah è vero // adesso vediamo // buon pranzo // a domani //	grazie	ciao emi // ciao // ciao // b
11	☐	ⓘ BOA3004	sembrava un pezzo di bacon // no no no xxx // // vuoi un pezzetto? // no	grazie	// cos' hai mangiato tu? //

**Grazie!**



**Finanziato
dall'Unione europea**
NextGenerationEU



**Ministero
dell'Università
e della Ricerca**



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

www.unibo.it