



Finanziato
dall'Unione europea
NextGenerationEU



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

XVI Convegno ASLI

Documentare la diversità linguistica: una nuova risorsa per l'italiano parlato da persone con esperienza di migrazione internazionale

*Silvia Ballarè, Eugenio Gorìa,
Caterina Mauri, Eleonora Zucchini*

13 settembre 2024 - Trieste

** The research leading to these results has received funding from Project "DiverSIta-Diversity in spoken Italian", prot. P2022RFR8T, CUP J53D23017320001, funded by EU in NextGenerationEU plan through the Italian "Bando Prin 2022 - D.D. 1409 del 14-09-2022"*

Panoramica

- ◆ Il corpus KIParla
- ◆ I nuovi moduli Stra-Parla: obiettivi, metodologia e sfide
- ◆ Alcuni dati sulla nuova raccolta
- ◆ Esempi dai dati
- ◆ Prossimi passi



Panoramica

- ◆ **Il corpus KIParla**
- ◆ I nuovi moduli Stra-Parla: obiettivi, metodologia e sfide
- ◆ Alcuni dati sulla nuova raccolta
- ◆ Esempi dai dati
- ◆ Prossimi passi



[Home](#) [Chi siamo](#) [Il corpus](#) [Moduli](#) [Cerca nel corpus](#)



www.kiparla.it



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Il corpus KIParla

- ✓ **Liberamente accessibile**
trascrizioni allineate ai file audio
- ✓ **Ricco corredo di metadati**
riguardo ai parlanti e alle conversazioni
- ✓ **Interfaccia di ricerca secondo standard internazionali**
NoSketchEngine, funzioni di ricerca avanzate: KWIC, liste di frequenza, etc.
- ✓ **Struttura modulare, incrementale e replicabile**
possibilità di espandere il corpus nel tempo

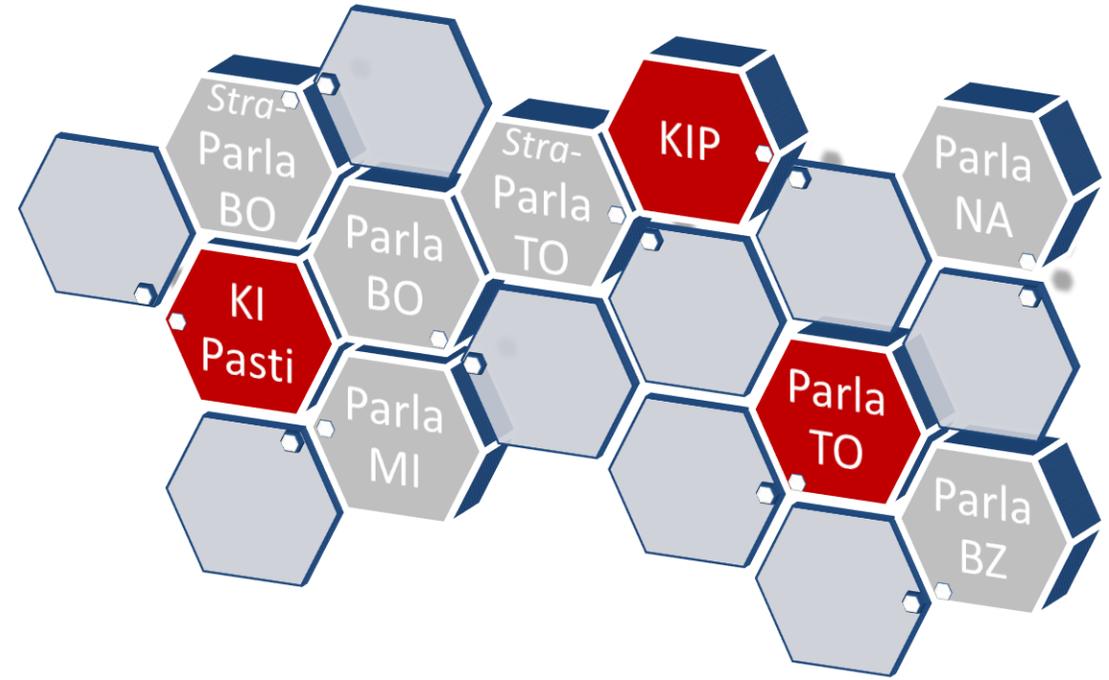


www.kiparla.it

(Mauri et al. 2019)

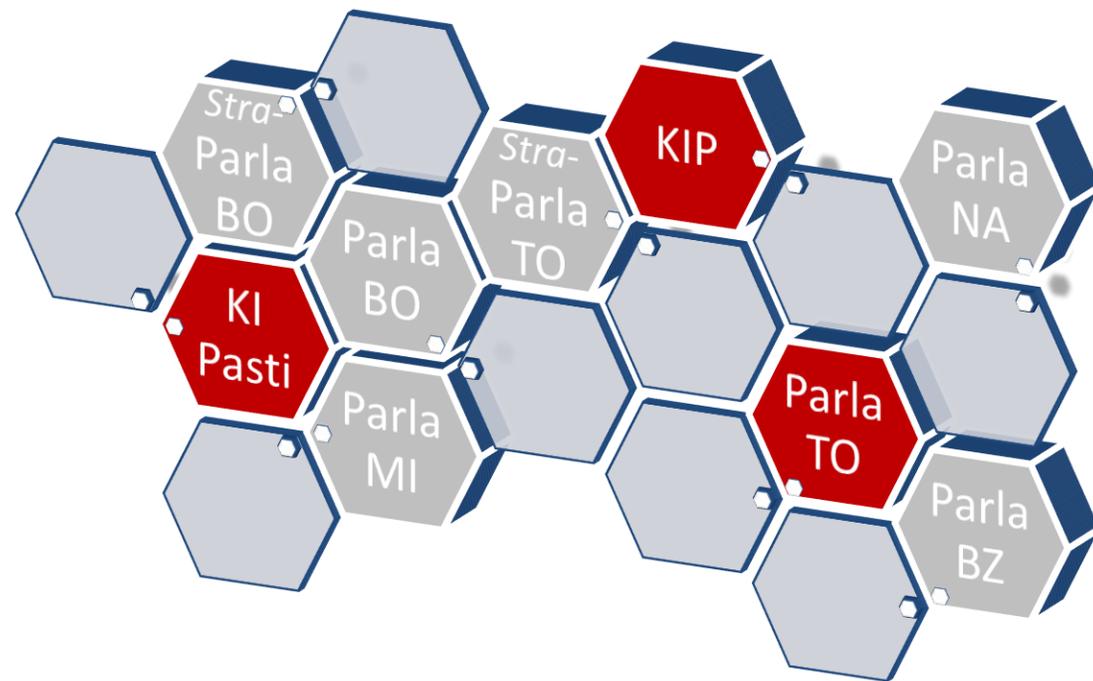
Struttura modulare e incrementale

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



Struttura modulare e incrementale

- **Modulare:** divisione interna in moduli indipendenti
 - ✓ Possibilità di consultare i moduli separatamente o in maniera congiunta

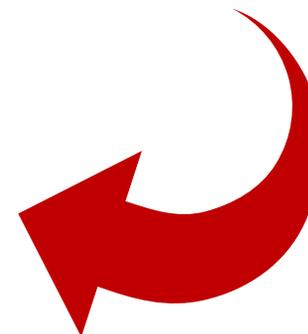


SELECT CORPUS



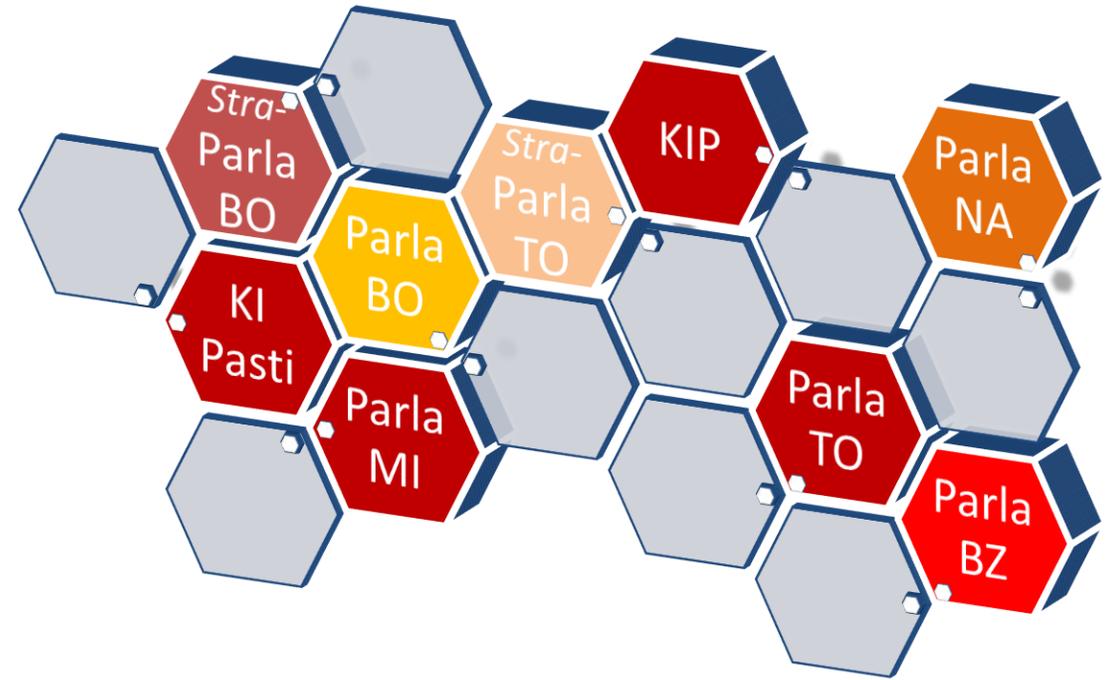
FREE CORPORA

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



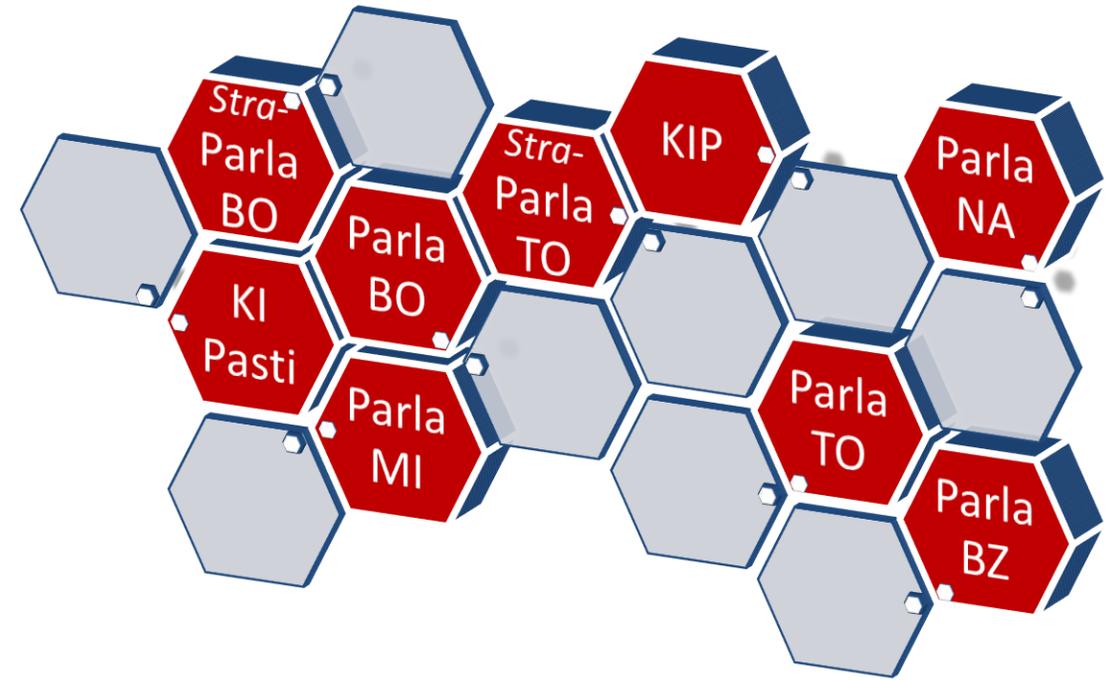
Struttura modulare e incrementale

- **Modulare:** divisione interna in moduli indipendenti
- **Incrementale:** possibilità di aggiungere nuovi moduli

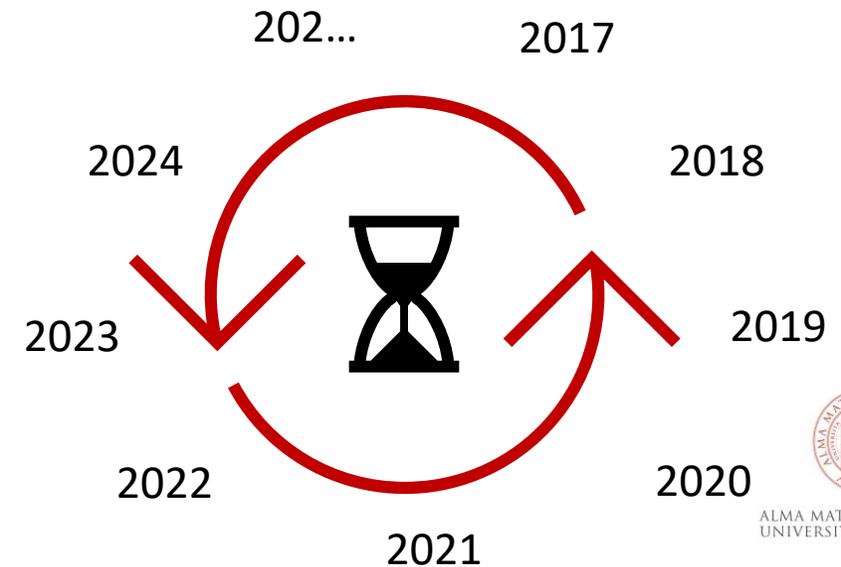


Struttura modulare e incrementale

- **Modulare:** divisione interna in moduli indipendenti
- **Incrementale:** possibilità di aggiungere nuovi moduli



Questa natura dinamica rende il corpus KIParla adatto a documentare la lingua parlata nel tempo.



Raccolta e trascrizione: *it takes a village!*



GDPR - Regolamento 2016/679

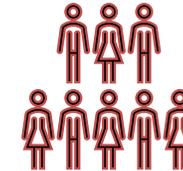
2018 - oggi: più di 100 studenti (Università di Bologna e Torino) hanno partecipato alla costruzione del corpus KIParla grazie a tirocini e tesi triennali e magistrali.



Formazione e supervisione



Aggiornamenti frequenti e incontri settimanali



Organizzazione e coordinamento

- **Trascrizioni** su ELAN, Jefferson (2004) semplificato per i tratti del parlato
- **Pseudoanonimizzazione:** rimozione dei dati sensibili dall'audio e dalla trascrizione.

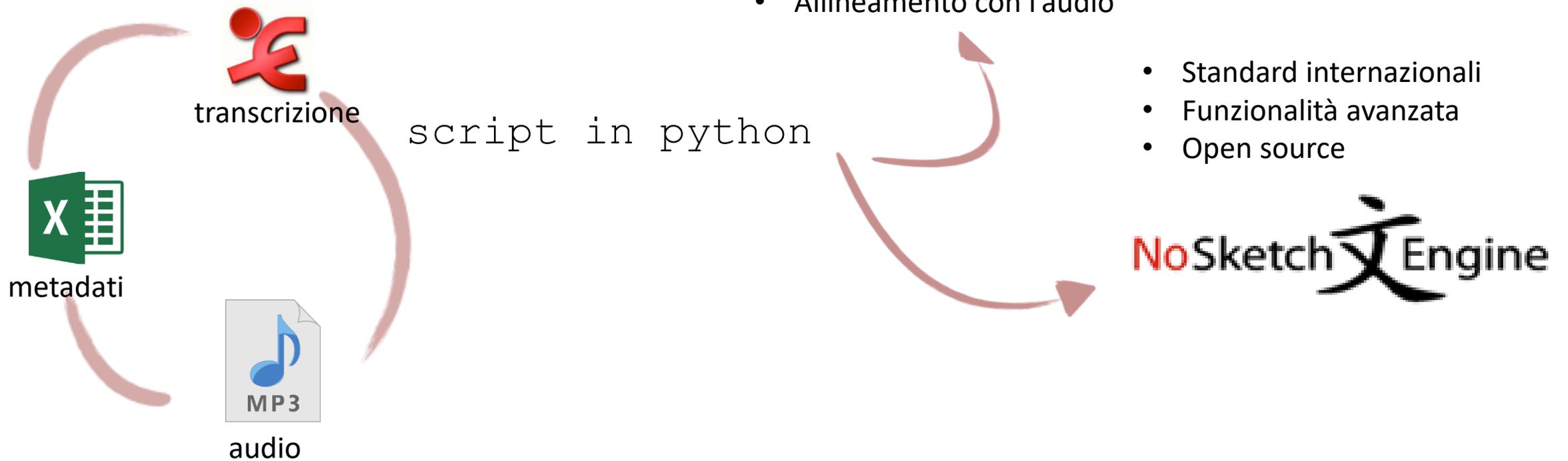


,	Intonazione ascendente	<ciao>	Rallentamento
.	Intonazione discendente	[hello]	Sovrapposizioni
:	Allungamenti	(hello)	Ipotesi del trascrittore
(.)	Pausa breve	xxx	Porzioni inintelligibile
>ciao<	Aumento velocità	((laughs))	Comportamento non verbale



Publicazione dei dati

- XML
- Txt (versione ortografica)
- Txt (versione conversazione)
- Associazione occorrenze/metadati
- Allineamento con l'audio



- Standard internazionali
- Funzionalità avanzata
- Open source



CONCORDANCE

KIPARLA



simple **ciao** • 299
184.03 per million tokens • 0.018%

Details

Left context KWIC Right context

1	<input type="checkbox"/>	TOD2015	ah ma quindi tu hai sentito cose che ha fatto jack ah okay // sì sì sì sì sì	ciao	jack se mai sentirai questo audio // // anche perchè camera mia e di jac	
2	<input type="checkbox"/>	PTB025	perché io ho due ore non posso perdere un'ora e quarantacinque e poi	ciao	ti saluto // // beh poi dipende dal film // quindi // mh // sì sì dipende d~ et	
3	<input type="checkbox"/>	KPN009	o poco dolci // ma insomma non hanno un po' sanno un po' di di miele //	ciao	matteo // cos'è? no no // sanno proprio del sapore quello verde della pa	
4	<input type="checkbox"/>	KPN022	amente andare dal campetto quello di cemento dove giocano a calcio //	ciao	marco grazie ciao // ehm perchè c'erano vari indizi e noi dovevamo indc	
5	<input type="checkbox"/>	KPN022	campetto quello di cemento dove giocano a calcio // ciao marco grazie	ciao	// ehm perchè c'erano vari indizi e noi dovevamo indovinare quale fosse	
6	<input type="checkbox"/>	KPN022	he resto qua // pesante // ti credo io che non l'hai mangiata tutta // okay	ciao	grazie ciao ciao // era intera // sì // come doveva essere // beh xxx tutte	
7	<input type="checkbox"/>	KPN022	a // pesante // ti credo io che non l'hai mangiata tutta // okay ciao grazie	ciao	ciao // era intera // sì // come doveva essere // beh xxx tutte pure patate	
8	<input type="checkbox"/>	KPN022	esante // ti credo io che non l'hai mangiata tutta // okay ciao grazie ciao	ciao	// era intera // sì // come doveva essere // beh xxx tutte pure patate // e	
9	<input type="checkbox"/>	KPN022	' perchè // che anni era // che anno era // è andato male // no // ah ok //	ciao	pluto // boh penso fosse tipo per i quat~ quindic'anni // boh // pluto // av	
10	<input type="checkbox"/>	KPN022	come che andava come // è la mirella // perchè la mirella // pronto // hei	ciao	// #ma cossa gaea el vivavoce // eh eh no penso che sia tipo il cellulare	
11	<input type="checkbox"/>	KPN022	il mio compleanno // ci sentiamo auguroni intanto // io non sarei felice //	ciao	ciao ciao ciao grazie ciao ciao ciao // perchè // mh io sono più vicino ai	
12	<input type="checkbox"/>	KPN022	o compleanno // ci sentiamo auguroni intanto // io non sarei felice // ciao	ciao	ciao ciao grazie ciao ciao ciao // perchè // mh io sono più vicino ai venti	
13	<input type="checkbox"/>	KPN022	pleanno // ci sentiamo auguroni intanto // io non sarei felice // ciao ciao	ciao	ciao grazie ciao ciao ciao // perchè // mh io sono più vicino ai ventiquatt	



ACCESSIBILITÀ

TRASCRIZIONI (ortografiche e conversazionali)

➤ Liberamente accessibili

AUDIO

➤ Liberamente accessibili previa registrazione sul sito

TOD2015

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

Audio TOD2015

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

8:26 / 30:52

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

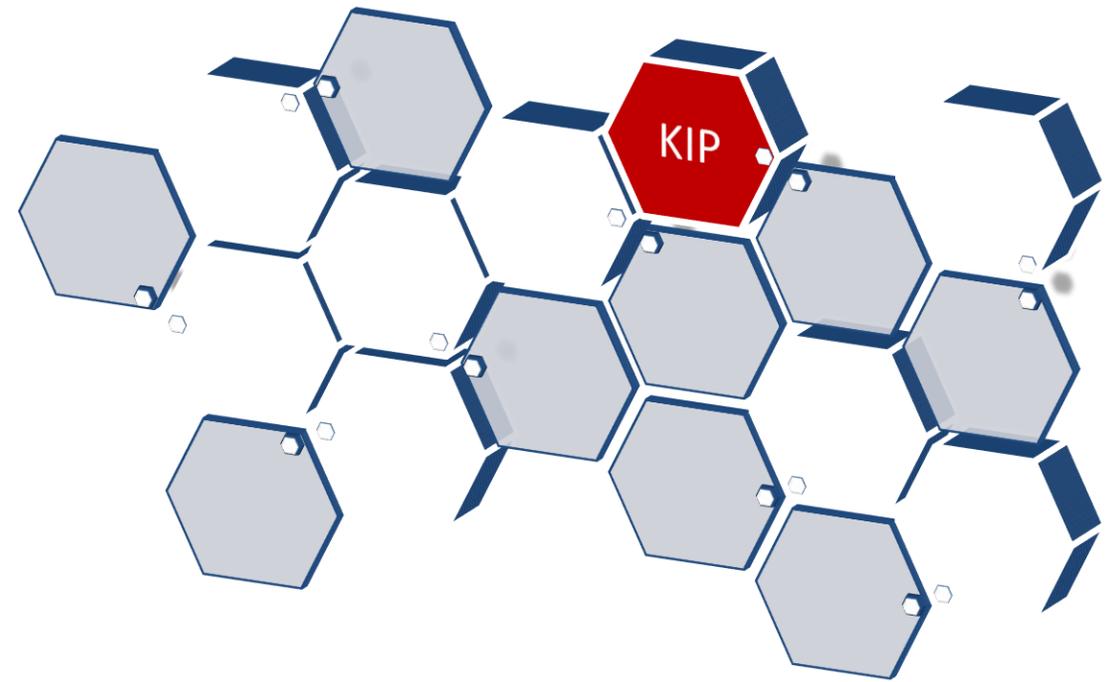
TO071	ti chiudevi in camera
TO073	esatto ma dall' altra parte
TO073	>si che poi era (.) attaccata a quella di jack< quindi era come se fossi in camera insieme in realtà ((ride))
TO071	ah ma quindi tu hai sentito cose [che ha fatto jack, ah okay]
TO073	[si si si (.) si si (.)] ciao jack se mai sentirai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

Modulo KIP

- 273 **parlanti colti** registrati in ambiente universitario

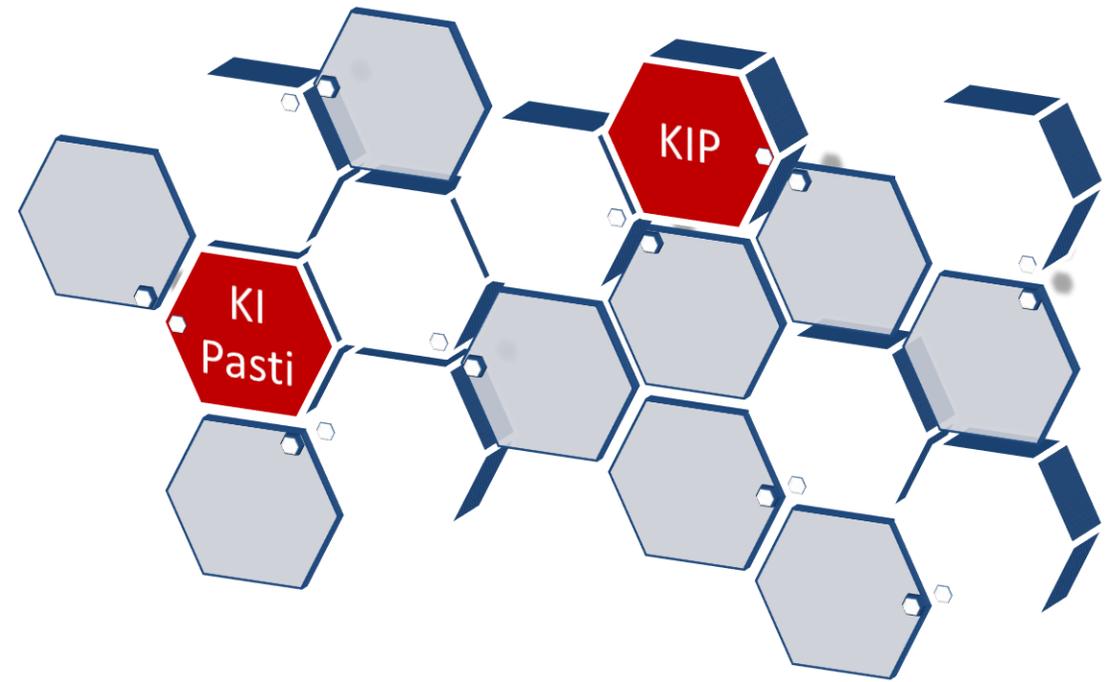
→ Osservare la variazione diafasica e diatopica

- ✓ Dati raccolti a Bologna and Torino (2016-2019)
- ✓ **Contesti comunicativi diversi**: 121 interazioni
- ✓ **TOT: ca. 70 hrs, 660,000 tokens**



Modulo KIPasti

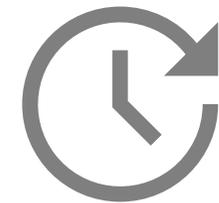
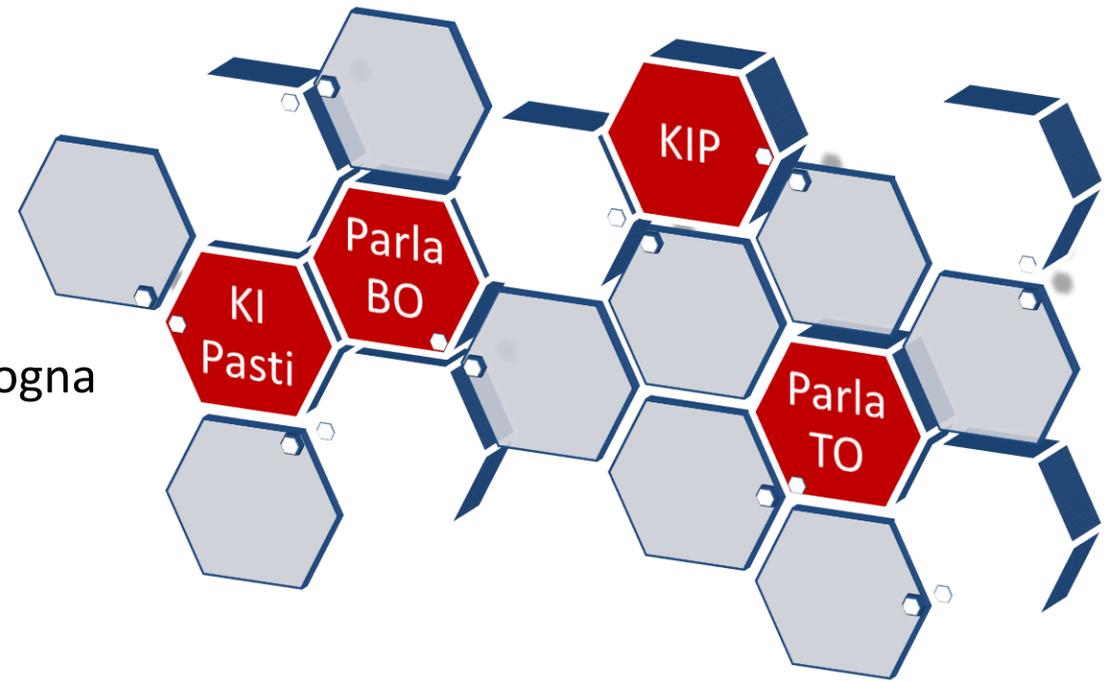
- 63 conversazioni a tavola raccolte in tutta Italia
- Osservare la variazione diatopica e diastratica e le varietà colloquiali.
- ✓ Dati raccolti in 13 regioni d'Italia (2020-2024)
 - ✓ 147 parlanti con **caratterizzazione sociale differente** (età, istruzione, occupazione)
 - ✓ Un contesto: **conversazioni a tavola**, dunque background comune, registro informale
 - ✓ **TOT: ca. 42 hrs, 487,000 tokens**



Moduli ParlaTO e ParlaBO

- Italiano parlato nelle città metropolitane di Torino e Bologna
- Osservare la variazione diatopica e diastratica e le varietà colloquiali.

- ✓ Dati raccolti a Torino (2019) e Bologna (2022-2024)
- ✓ 88 + 158 parlanti con **caratterizzazione sociale differente** (età, istruzione, occupazione)
- ✓ Un contesto: 67 + 86 interviste semi-strutturate
- ✓ **TOT: ca. 50 + 66 hrs, 560,000 + 650,000 tokens**



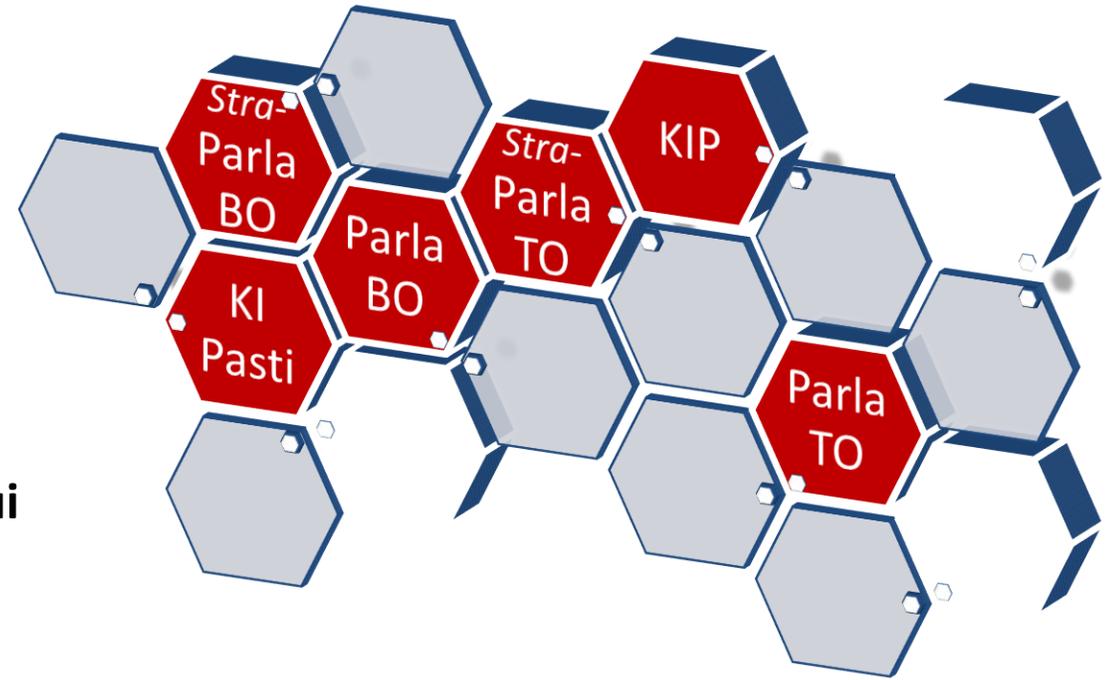
TO BE PUBLISHED
VERY SOON



Moduli Stra-ParlaBO e Stra-ParlaTO

Prossimi passi: **Moduli Stra-ParlaBO e Stra-ParlaTO**

- Italiano parlato all'interno di comunità di **parlanti con origine straniera con repertori multilingui**



→ Osservare le varietà di apprendimento e la variazione intergenerazionale e diastratica

- ✓ Dati raccolti a Bologna e Torino (2024 – in corso)
- ✓ Parlanti con **repertori multilingui, caratterizzazione sociale e origine**
- ✓ **EXPECTED: ca. 128 hrs, 1,250,000 tokens**

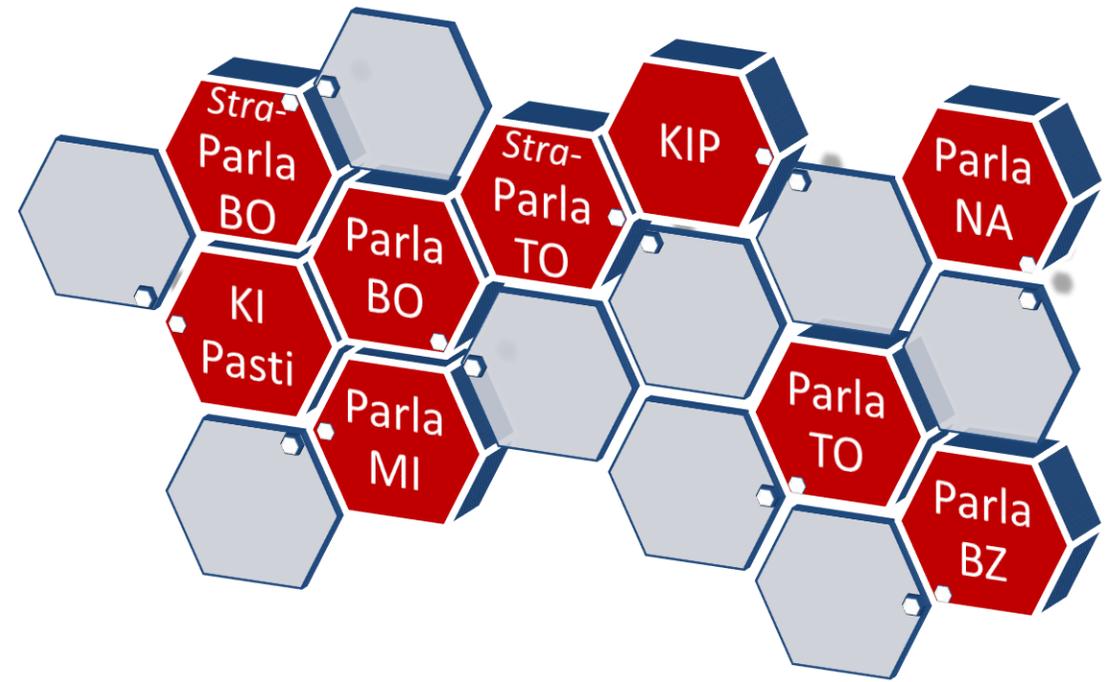
Raccolta dati in corso!



Collaborazioni in corso

Prossimi passi: **moduli ParlaBZ, ParlaNA, ParlaMI**

- Italiano parlato a **Bolzano** (Daniela Veronesi),
- Italiano parlato a **Napoli** (Margherita di Salvo),
- Italiano parlato a **Milano** (Federica da Milano)



→ Osservare la varietà informale e la variazione regionale e diastratica.

Work in progress!



Panoramica

- ◆ Il corpus KIParla
- ◆ **I nuovi moduli Stra-Parla: obiettivi, metodologia e sfide**
- ◆ Alcuni dati sulla nuova raccolta
- ◆ Esempi dai dati
- ◆ Prossimi passi

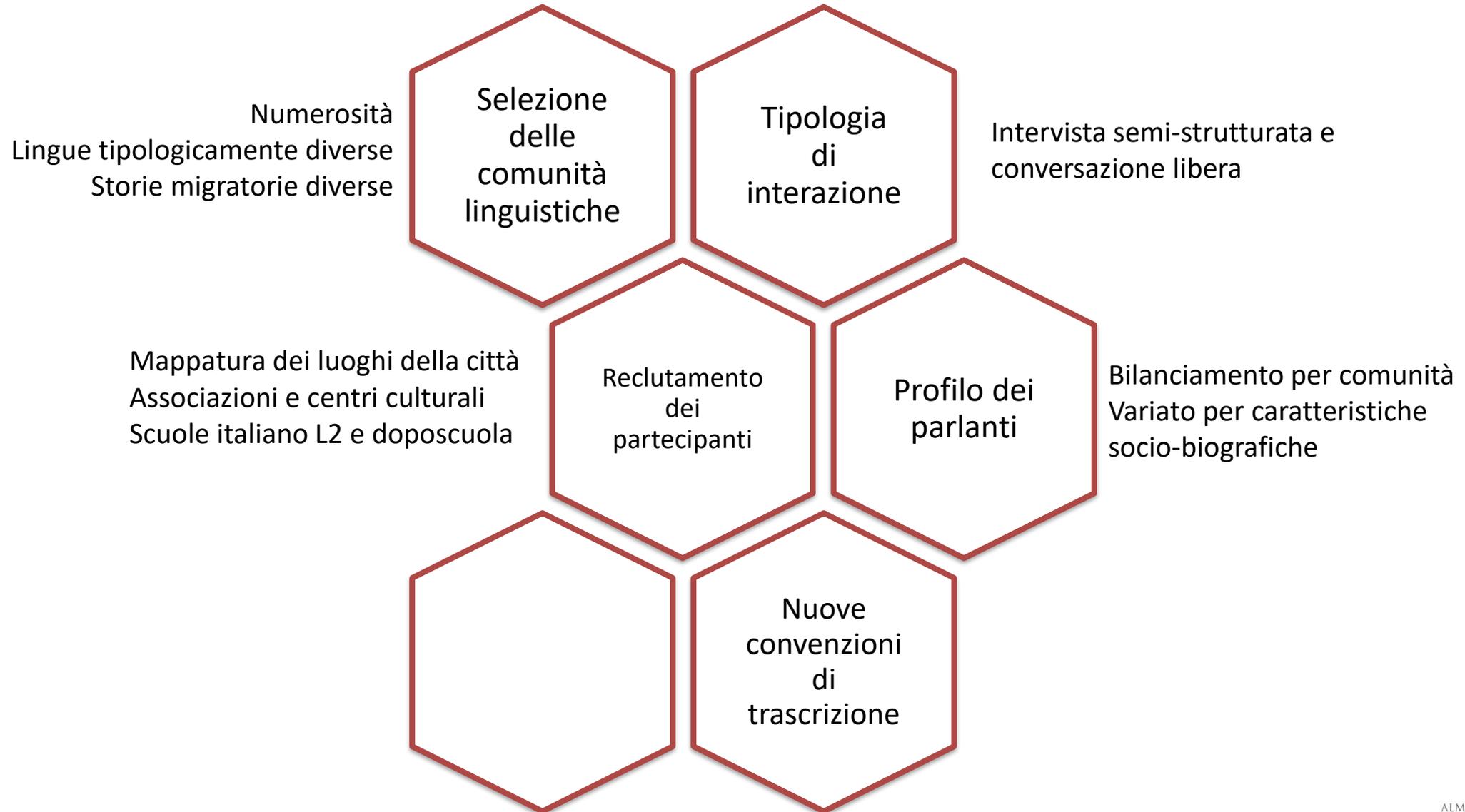


Obiettivi: perché i moduli stranieri

- ◆ Contesti urbani multiculturali
- ◆ Sguardo più ampio quando si studiano fenomeni linguistici
- ◆ Indagare ruolo di parlanti plurilingui e apprendenti nell'emergere di nuovi fenomeni e nuove costruzioni
- ◆ Confrontare varietà di apprendimento con altre varietà (punti in comune e differenze)



Moduli Stra-Parla: metodologia



Quali lingue e quali comunità

BOLOGNA: tot 64 h

UCRAINO



ARABO MAROCCHINO



BENGALESE



CINESE



RUMENO



SPAGNOLO PERUVIANO



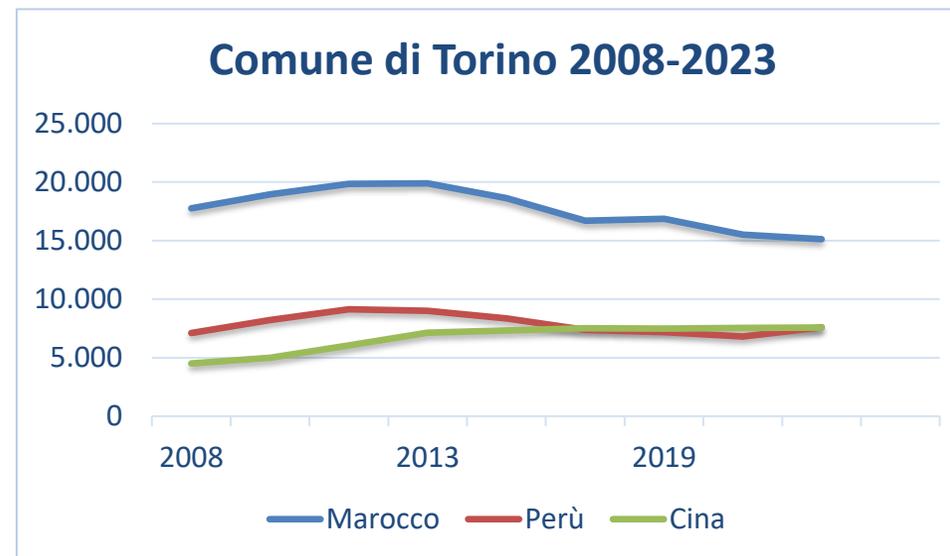
tot
Stra-Parla
128 h

TORINO: tot 64 h

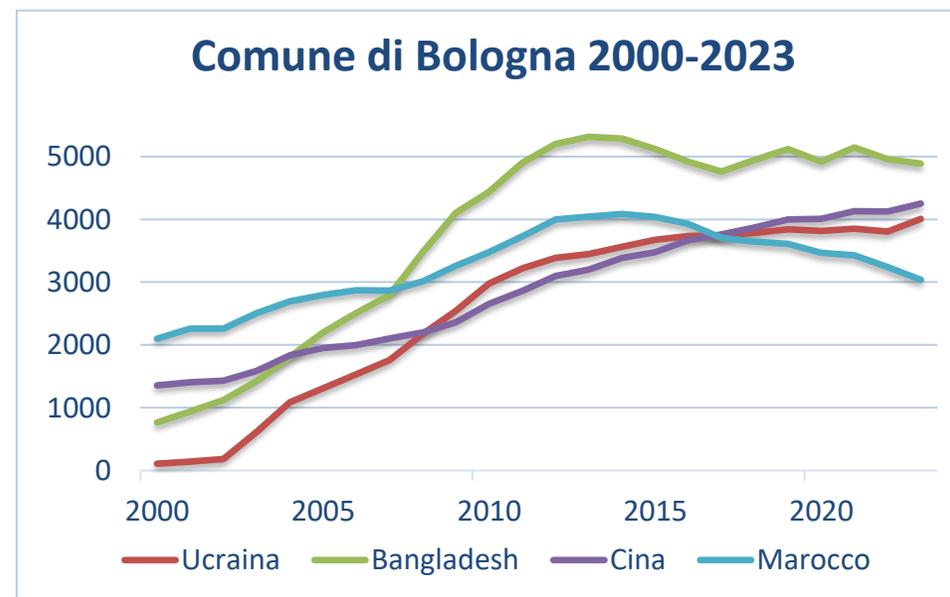


Migrazioni vecchie e nuove

Anno	Romania	Marocco	Perù	Cina
2008	47.771	17.773	7.104	4.506
2023	45.194	15.123	7.602	7.576



Anno	Ucraina	Bangladesh	Cina	Marocco
2000	108	756	1.356	2.100
2023	4890	3039	4.251	4.007



Dati da [Home | I numeri di Bologna \(inumeridibolognametropolitana.it\)](#)
[Benvenuti - aperTO \(comune.torino.it\)](#)





L'intervista semistrutturata

- Strumento di vecchia data e molto diffuso per la raccolta dati (Tagliamonte 2006)
- Macro-obiettivo: raccogliere dati di parlato informale
- Come?
 - ~ creare un clima informale, far sì che l'intervista sia un vero e proprio dialogo
 - ~ *elicit **narratives of personal experience**, where community norms and styles of personal interaction are most plainly revealed, and where style is regularly shifted towards the vernacular;*
 - ~ *stimulate **group interaction** among the people present, and so record conversation not addressed to the interviewer;*

(Labov 1984: 32-33)

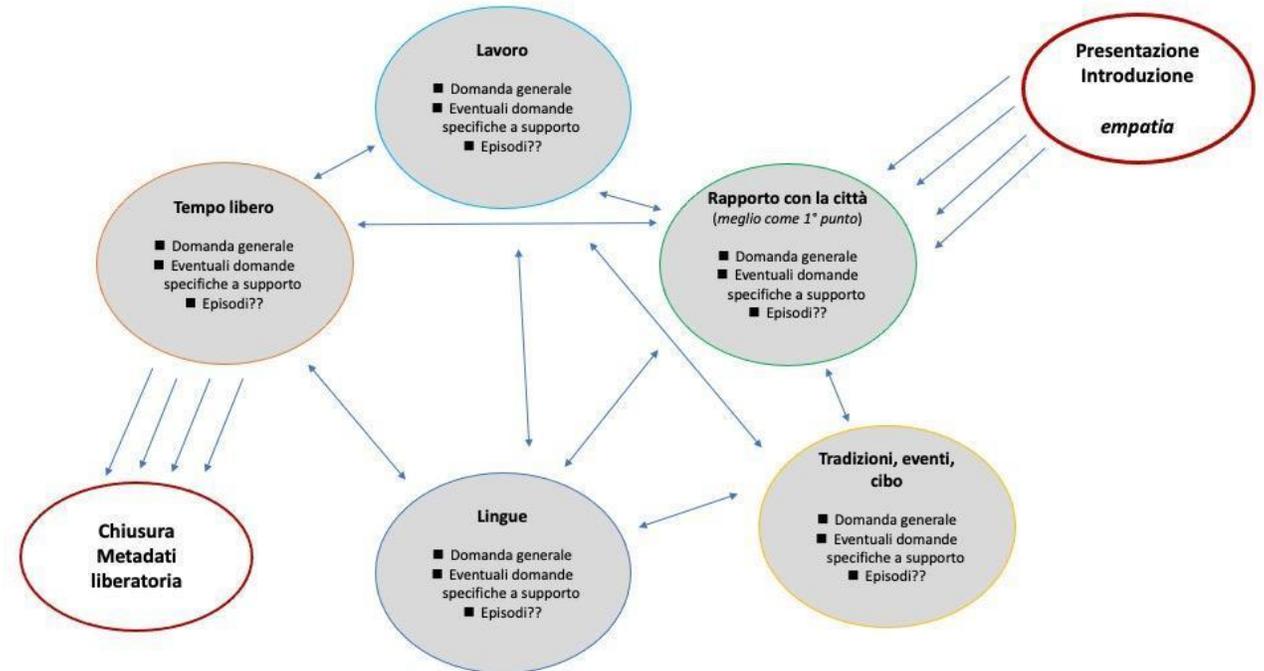




Come affrontare l'intervista

- Domande come spunti
- Ordine non prestabilito
- Libere associazioni fra gli argomenti
- *isolate from a range of **topics those of greatest interest** to the speaker, and allow him or her to lead in defining the topic of conversation.*

(Labov 1984)



ParlaTO e ParlaBO: l'intervista



Blocchi tematici:

- Rapporto con la città
- Tradizioni, eventi, cibo
- Lavoro e tempo libero
- Uso del dialetto



Stra-ParlaBO e Stra-ParlaTO: l'intervista



Blocchi tematici:

- Rapporto con la città
- **Tradizioni, eventi, cibo**
- Lavoro e tempo libero
- **Lingue**



- Qui in Italia si mantengono anche tradizioni del tuo paese di origine?
- Ti piace la cucina tradizionale di Bologna / Torino o preferisci mangiare piatti tipici del tuo paese di origine ?



- Che lingue conosci?
- In quali contesti le parli?
- Che rapporto hai con l'italiano?
- Come e dove l'hai imparato?
- Che rapporto hai con la lingua d'origine?



Intervista semistrutturata: turno esteso



(1) SBIU002

PSB010 perché le persone passavamo c'è una pista ciclabile lungo la: questa la ferrovia passano le persone anziane e vedevano la diversità di tutti i fiori cioè avevo fiori ho le rose che fioriscono (.) <a natale> c'era la neve e sotto i fiori quelle:: rose senza spine a me dispiace che va tutto: (.) °rovinato però adesso° troverò qualche soluzione



Metadati PSB010

Origine: Ucraina

Madrelingua: Ucraino

Fascia d'età: 51-55

Titolo di studio: diploma

Professione: art

Permanenza in Italia: >10 anni

Età all'arrivo: >18 anni

Intervista semi-strutturata: sequenza dialogica



(2) SBIA003

PSB017 in germania lo stesso per andare in germania
devi avere un un cert-
certificato attestazione

PSB005 mhmh

PSB017 #duc come si chiama #duc ehm?

PSB016 sì ma non centra

PSB017 no per me è questo x

PSB005 [cioè secondo te sarebbe meglio]

PSB017 [xx xx io:]

PSB016 [vabbè]

uno [se vuole imparare può studiare sempre[:]]

PSB017 [io: non:]

[questa è un'altra co]sa discorso

come esempio come

PSB016 ah sì [quelle lì ti fanno ti mettono più limite]

PSB017 [no io ho parlato per l'inizio] scusami allora



Metadati PSB016; PSB017

Origine: Morocco

Madrelingua: Arabo marocchino

Età: 36-40; 41-45

Titolo di studio: diploma

Professione: pnq

Permanenza in Italia: >10 anni

Età all'arrivo: >18 anni

Conversazione libera: esempio



(3) SBCU001

PSB022: nipote a casa
sempre de- devo preparare cosa
lui sempre NONNA vuoi q- questo nonna vuoi questo
°mamma°

PSB005: ma quanto è rimasto qui
quanto si è quanto tempo è stato qui

PSB022: e:h dieci giorni

PSB005: ah okay

PSB022: mhmh

PSB005: e che cosa avete fatto

PSB022: adesso a malta,

PSB005: okay [okay]

PSB022: <[poi] otto maggio fino a dic- (.) diciotto>
america

PSB005: °ah° ma dove?

PSB022: e:h l: los angeles

PSB005: ma in <vacanza>?

PSB022: ehm là mh
lui amica studia là già

PSB005: [mh okay okay]

PSB022: [questo no-] nove maggio finito e tutti
ami[ci qua]ttro ((ride)) xx

PSB005: [ah] (.) a:h cioè vanno tipo a fare una festa

PSB022: eh sì



Metadati PSB022

Origine: Ucraina

Madrelingua: Russo

Età: 71-75

Permanenza in Italia: >10 anni

Età di arrivo: >18 anni

Titolo di studio: diploma

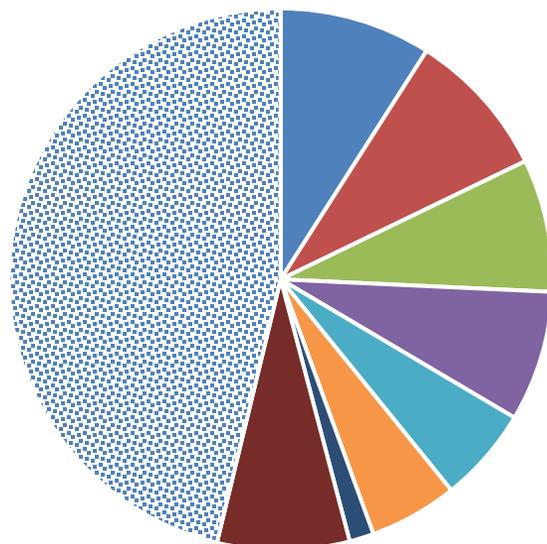
Professione: pnq

Panoramica

- ◆ Il corpus KIParla
- ◆ I nuovi moduli Stra-Parla: obiettivi, metodologia e sfide
- ◆ **Alcuni dati sulla nuova raccolta**
- ◆ Esempi dai dati
- ◆ Prossimi passi



Str-Parla: a che punto siamo?



■ C ■ U ■ A ■ B ■ P ■ M ■ C ■ R ■ Da raccogliere

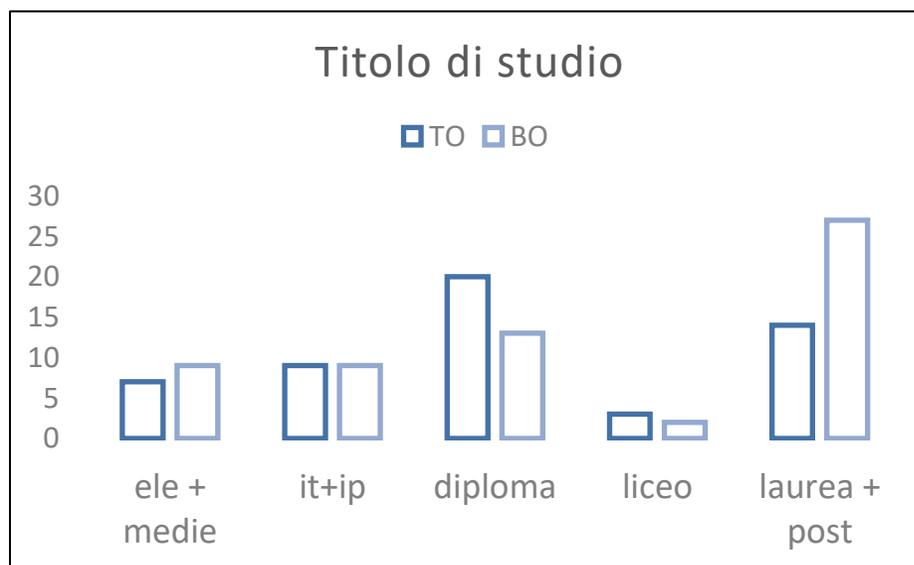
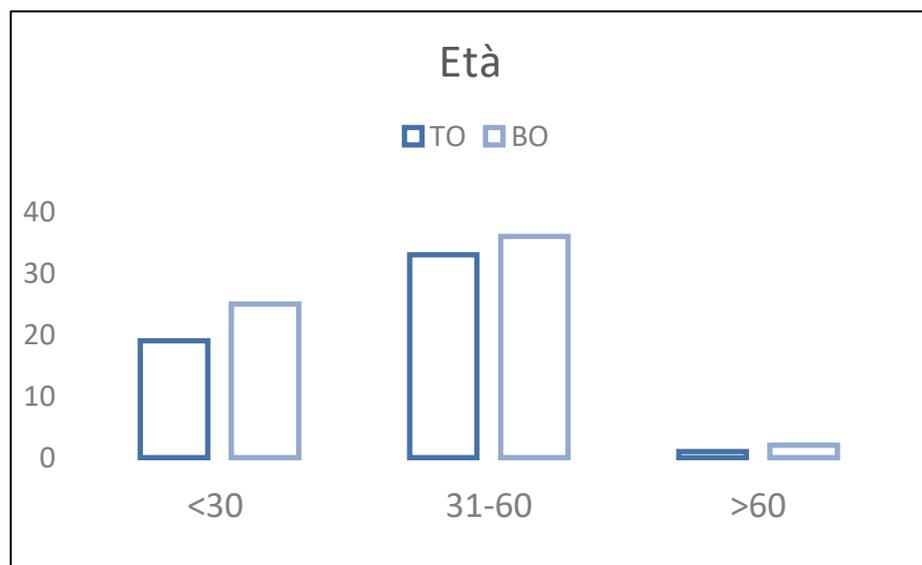
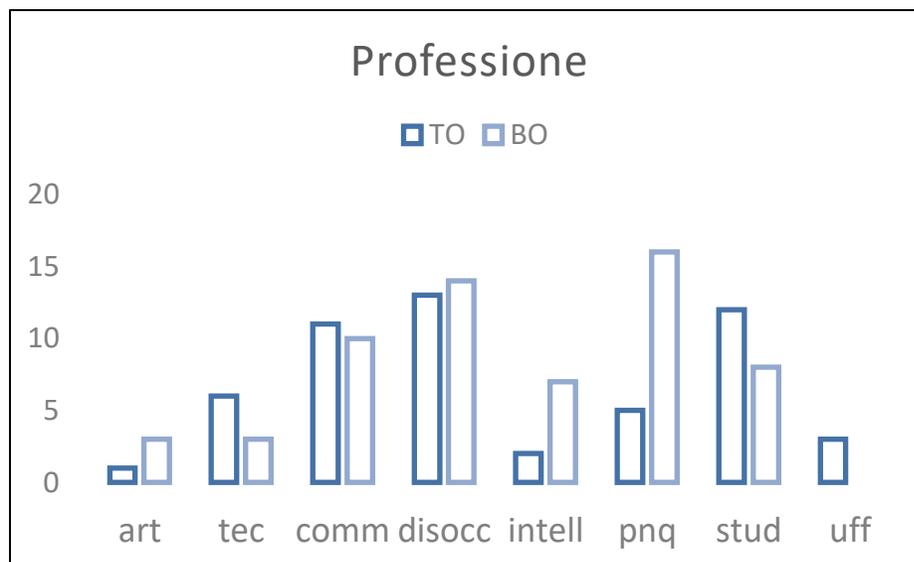
Comunità linguistica	Intervista	Conversazione libera	Totale
Cinese BO	9:19:00	2:13:39	11:32:39
Ucraina	9:46:27	1:29:03	11:15:30
Marocchina BO	8:32:34	1:35:54	10:08:28
Bengalese	9:58:16	0	9:58:16
Peruviana	06:28:15	00:47:18	07:15:33
Marocchina TO	06:34:43	00:09:17	06:44:00
Cinese TO	01:51:38	0	01:51:38
Romania	10:03:10	0	10:03:10
Totale	37:23:40	5:18:36	68:49:14



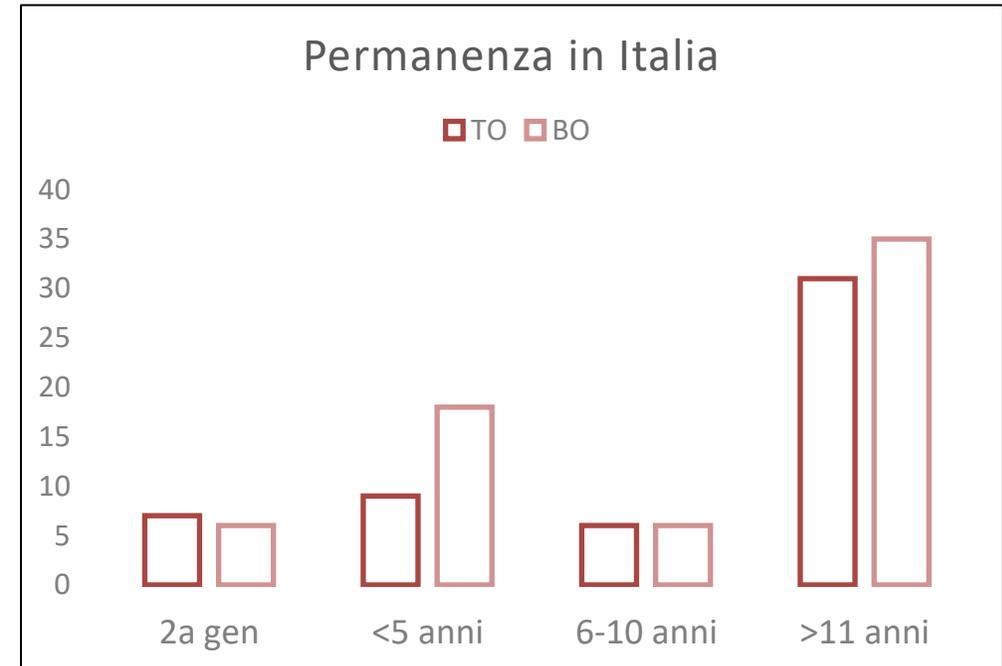
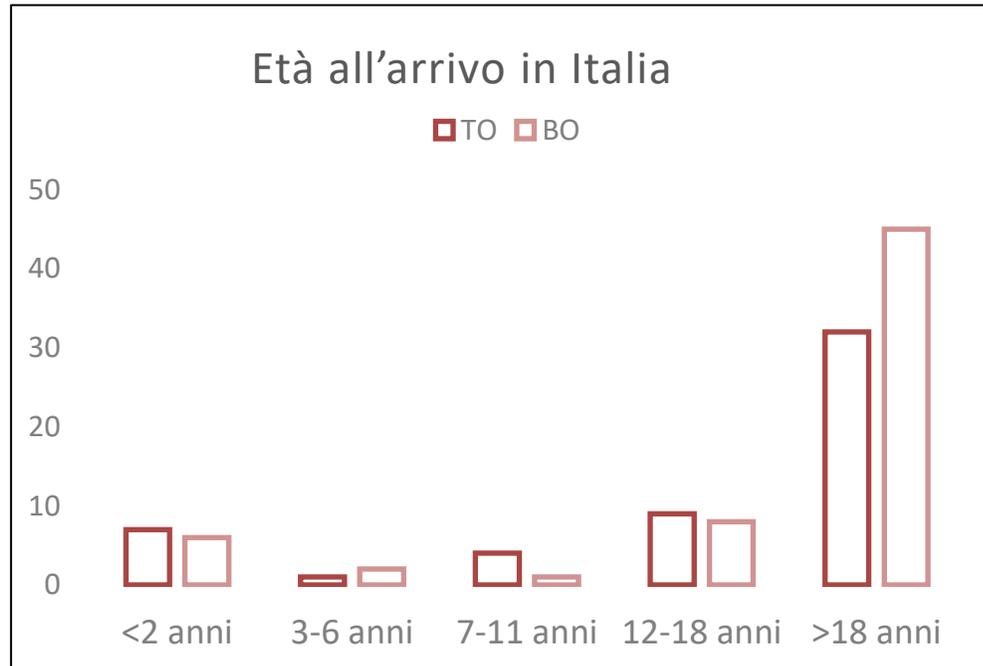
Profilo dei parlanti dei moduli StraParla



Tot. partecipanti: 120



Profilo dei parlanti dei moduli StraParla



Panoramica

- ◆ Il corpus KIParla
- ◆ I nuovi moduli Stra-Parla: obiettivi, metodologia e sfide
- ◆ Alcuni dati sulla nuova raccolta
- ◆ **Esempi dai dati**
- ◆ Prossimi passi



Esempi dai dati: varietà iniziali



(4) SBIB009

PSB058: cercare **lavorare** difficile

PSB005: [qui] (.) o in bangladesh?

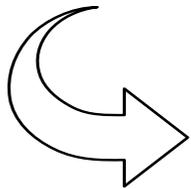
PSB058: [sì] no: bangladesh tanti **lavorare** // **tante lavoro** c'è

PSB005: qui è più difficile?

PSB058: sì sì cercare difficile **lavorare**

però bangladesh lavorare (.) soldi di meno

➤ **Comparabilità con altre varietà di italiano**
e.g. it. popolare; it. informale



- ✓ Organizzazione della sintassi e ordine delle parole
- ✓ PoS
- ✓ Costruzioni presentative e esistenziali
- ✓ Morfologia e uso delle forme verbali
- ✓ Assegnazione e accordo di genere

Metadati PSB058

Comunità linguistica: bengalese

Madrelingua: bengalese

Fascia d'età: 16-20

Permanenza in Italia : <5 anni

Titolo di studio: medie (ip in corso)

Professione: studente ip

Esempi dai dati: varietà iniziali



(5) SBIC005

PSB009: questa estate la mia accademia con eh accademia di firenze // c'è un: qualche attività

PSB005: sì [sì sì]

PSB009: [xx] // sono arrivato a- arrivata a firenze e studio circa: (quindi) giorni

PSB005: okay

PSB009: e:hm // e un mo- è un periodo molto: (.) bello per me // m:h

è è xx speciale eh

il professore pa- pa- parla (delle case) e poi // ah questo pittura è nel- nel chia- chiesa vicina

quindi: andiamo a vede[re]

PSB005: [mh]mh

PSB009: oh

non non solo ve- gua- guardiamo un foto su solo sul: e:hm ppt

PSB005: [mhmh]

PSB009: [opp]ure sul compu[ter lo] possiamo: ve-

PSB005: [okay sì]

PSB009: [da oc-]

PSB005: [vedere] sì

PSB009: [ve]dere da: occhi?



Metadati PSB009

Comunità linguistica: cinese

Fascia d'età: 21-25

Titolo di studio: laurea

Professione: intell

Permanenza in Italia: <5 anni

Età all'arrivo: >18

Esempi dai dati: varietà avanzate



(6) SBIB003

PSB029: e se tu parli se tu chiedi una persona da anziana
che è finito studiare cioè

come straniero come me se come straniero significa che sei venuto da un'altra città

PSB005: mhmh

PSB029: tipo anche napoli eh giù parliamo anche ROMA

[gua]rda che se chiedi che quanti anni sei in: sei in bologna tra parentesi

PSB005: [mhmh]

PSB029: dici guarda che io ho ho laureato qui

PSB005: mhmh

PSB029: e ti racconta la storia come era prima

e trattava come: come straniero cioè non te- non dà la c[asa] non c'è proprio

PSB005: [xxx]

PSB029: era il bologna ha questa storia [da:] da storia [cioè proprio] da cento anni mi sem[brava ancora più ((ride)) sì]

Metadati PSB029

Comunità linguistica: bengalese

Fascia d'età: 31-35

Titolo di studio: ip

Professione: comm

Permanenza in Italia: >10

Età all'arrivo: 12-18



Esempi dai dati: code-switching (intervista) I



(7) 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: okay

well, eight ways

eight ways, i thought you meant eight cuisines

yes, cuisines I wanted to say

but cuisines I don't know how to translate,

eight ways, maybe you can't say it

- Code switching legato ai partecipanti e al discorso
- *Focus on form*
- Figura del mediatore

Metadati PSB001; PSB002

Comunità linguistica: cinese

Madrelingua: cinese mandarino; dialetto xiang

Fascia d'età: 26-30; 31-35

Titolo di studio: PhD; laurea

Professione: intell; stud

Permanenza in Italia: <5 anni

Età all'arrivo: >18 anni

Esempi dai dati: code-switching (intervista) II



(8) SBIA013

PSB005: qual è la più bella la città più bella del marocco

PSB069: no no agadir

PSB005: okay

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 ((ride))

PSB066: ((ride))

PSB067: no no la più bella è casablanca

PSB005: okay



- Code-switching legato ai partecipanti

Agadir is nice?

I have never been there from Casablanca

do you like Agadir? Is it better than Casablanca? not at all

I am getting angry ((laughs))

Metadati PSB067; PSB068; PS069

Comunità linguistica: marocchina
Madrelingua: marocchino;
francese

Fascia d'età: 41-45; 46-50

Titolo di studio: diploma

Professioni: pnq; disocc

Permanenza in Italia: >10 anni

Età all'arrivo: > 18 anni

Esempi dai dati: code-switching (conversazione libera)



- Code-switching legato ai partecipanti

Metadati PSB043; PSB045

Comunità linguistica: marocchina

Madrelingua: marocchino; francese

Fascia d'età: 21-25; 41-45

Titolo di studio: lic; medie

Professioni: stud; disocc

Permanenza in Italia: >10 anni

Età all'arrivo: 3 anni; > 18 anni

Panoramica

- ◆ Il corpus KIParla
- ◆ I nuovi moduli Stra-Parla: obiettivi, metodologia e sfide
- ◆ Alcuni dati sulla nuova raccolta
- ◆ Esempi dai dati
- ◆ **Prossimi passi**



Linee guida per una risorsa Modulare Incrementali e Replicabile

KIParla è la prima risorsa Modulare, Incrementale e Replicabile per l'italiano parlato (**risorsa MIR**)

- ✓ Tutte le fasi della costruzione della risorsa sono documentate, in accordo con i principi FAIR (Wilkinson et al. 2016)
- **Linee guida MIR: tutto il protocollo sarà accessibile e replicabile.**



Sviluppi e sfide future

Sviluppi computazionali:

- ✓ Trascrizione semi-automatica (?!?)
- ✓ Lemmatizzazione
- ✓ PoS tagging
- ✓ UD Treebank dell'italiano parlato, basata sul KIParla

BOSCO, Cristina et al. 2020.

KIPoS @ EVALITA2020: Overview of the Task on KIParla Part of Speech Tagging In: *EVALITA Evaluation of NLP and Speech Tools for Italian - December 17th, 2020: Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian Final Workshop* [online]. Turin: Accademia University Press, 2020. Available on the Internet: <[http:// books.openedition.org/aaccademia/7743](http://books.openedition.org/aaccademia/7743)>. ISBN: 9791280136329. DOI: [https://doi.org/10.4000/ books.aaccademia.7743](https://doi.org/10.4000/books.aaccademia.7743).

Espansione del corpus:

- ✓ **Nuovi moduli** >> nuove collaborazioni





simple grazie • 772

475.15 per million tokens • 0.048%



Details

Left context

KWIC

Right context

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





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