



Ministero
dell'Università
e della Ricerca



Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DiverSIta – Diversity in Spoken Italian

CLUB_Progetti

03/04/2024

Caterina Mauri, Eugenio Gorla, Eleonora Zucchini

PRIN 2022 PNRR - DiverSIta 'Diversity in Spoken Italian'

- Finanziamento MUR, programma PRIN 2022 PNRR, n. P2022RFR8T
- Dal 30/11/23 al 30/11/25
- Due unità di ricerca: **Università di Bologna** e **Università di Torino**
- PI: Caterina Mauri

<https://site.unibo.it/divers-ita/en>

UR Bologna: Caterina Mauri, Claudia Borghetti, Nicola Grandi, Yahis Martari, Rosa Pugliese, Eleonora Zucchini & Silvia Ballarè, Ludovica Pannitto, Matteo Pascoli & tirocinanti KIParla

UR Torino: Eugenio Goria, Cecilia M. Andorno, Beatrice Bernasconi, Massimo Cerruti, Paolo Della Putta, Guglielmo Inglese & tirocinanti KIParla

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"



PRIN 2022 PNRR - DiverSIta '*Diversity in Spoken Italian*'

- Documentare il **patrimonio culturale orale dell'italiano**, nella sua dimensione dinamica
- ✓ particolare attenzione ai contesti urbani di **Bologna** e **Torino** - il linguaggio è il primo strumento per costruire esclusione o inclusione, gruppi o isolamenti

*Una migliore rappresentazione della diversità
insita nell'italiano parlato
è un passo cruciale verso una società più inclusiva,
in grado cioè di rappresentare
la diversità dei suoi individui.*



PRIN 2022 PNRR - DiverSIta 'Diversity in Spoken Italian'

Cos'è l'italiano? ...e chi lo parla?

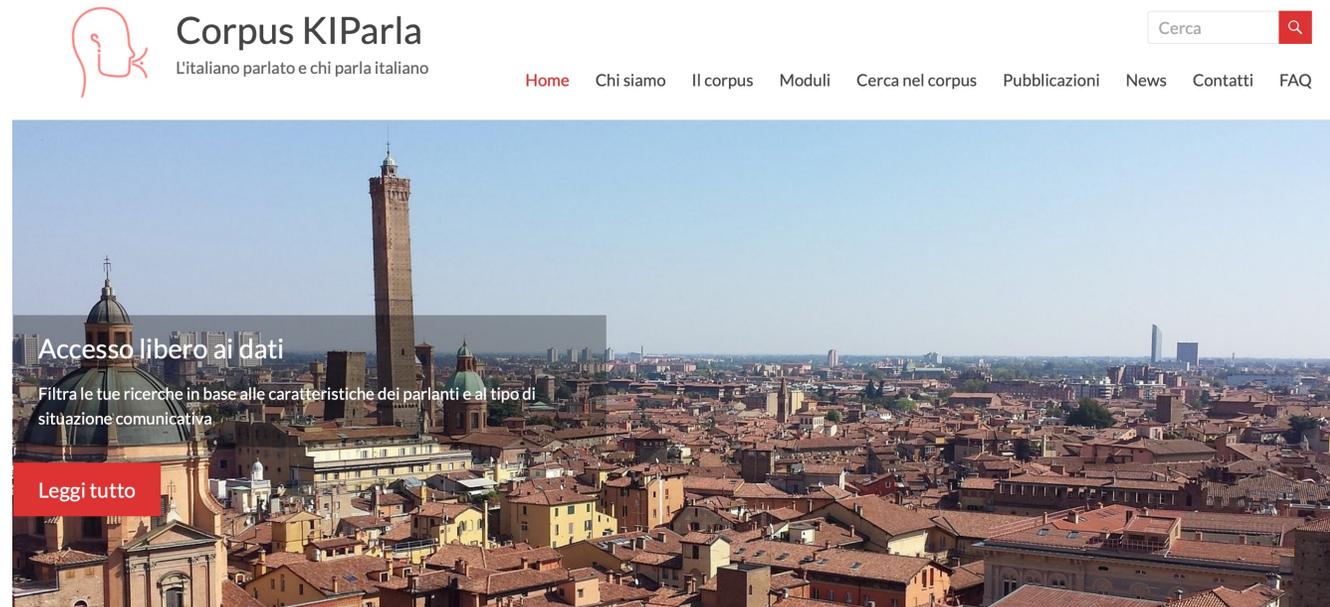
- Intendiamo rappresentare e comprendere:
 - ✓ la variegata **costellazione di parlanti di italiano**
 - ✓ le numerose **varietà di italiano** che sono parlate

Come?

... non partiamo da zero!

Espansione di una **risorsa digitale** esistente per lo studio dell'italiano parlato, il **KIParla** www.kiparla.it:

- campioni di dati bilanciati e anonimizzati;
- audio e trascrizioni ad accesso libero;
- interrogazione tramite NoSketchEngine



Corpus KIParla
L'italiano parlato e chi parla italiano

Cerca

Home Chi siamo Il corpus Moduli Cerca nel corpus Pubblicazioni News Contatti FAQ

Accesso libero ai dati
Filtra le tue ricerche in base alle caratteristiche dei parlanti e al tipo di situazione comunicativa

Leggi tutto

PRIN 2022 PNRR - DiverSIta '*Diversity in Spoken Italian*'

- **Vari tipi di interazione** → Interviste semistrutturate, conversazioni libere, conversazioni a tavola, lezioni universitarie, esami, ricevimento studenti, ...
- **Variazione sociodemografica dei parlanti** → Giovani universitari e non, adulti lavoratori e non, anziani con il titolo di terza media o con il dottorato, diverse regioni di provenienza, diverse professioni, ...

KIParla oggi:

110 h totali, 340 parlanti
ca. 1.125.996 token

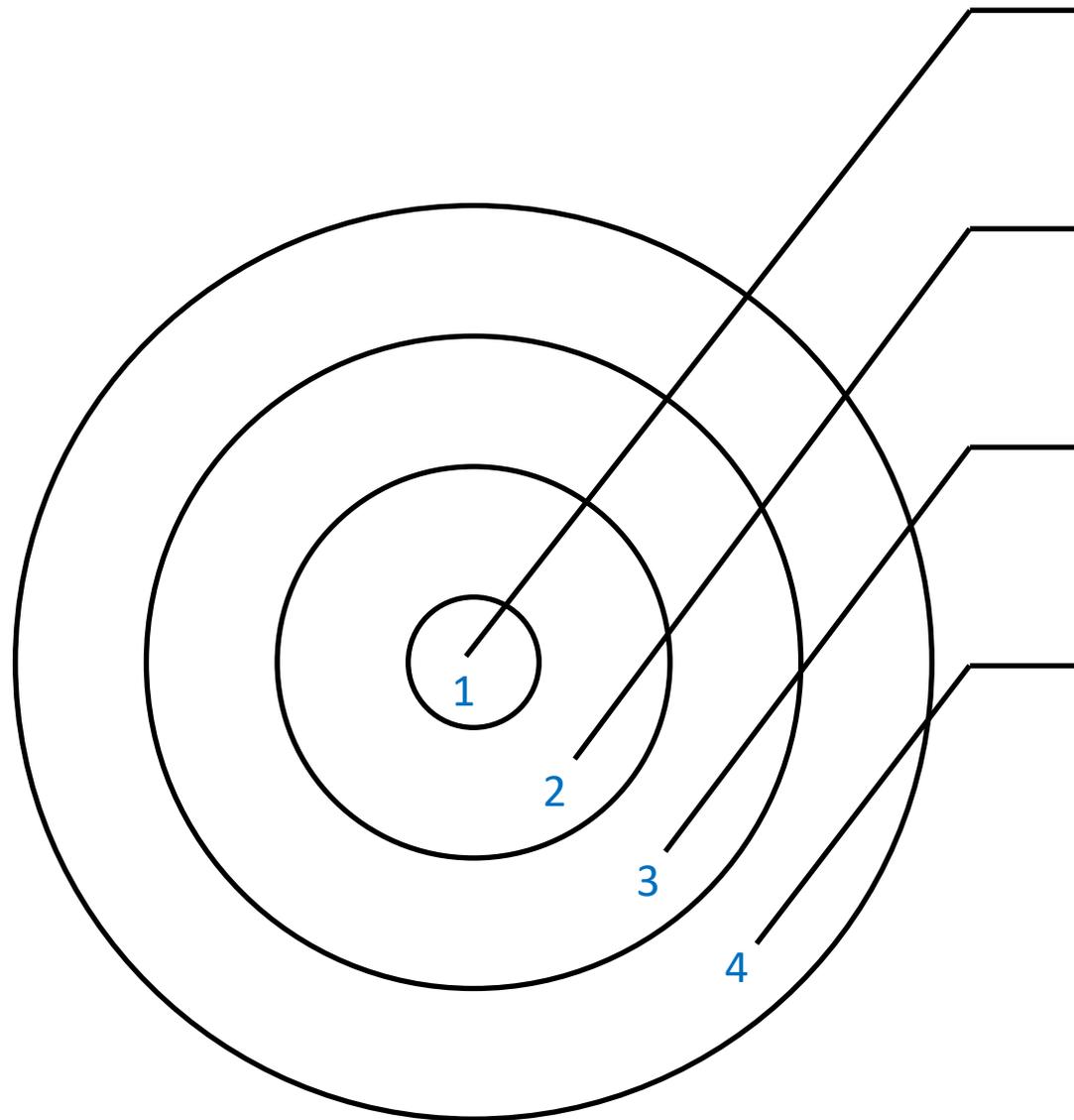


Obiettivo di DiverSIta:

340 h totali, ca. 900 parlanti
ca. 3.700.000 token



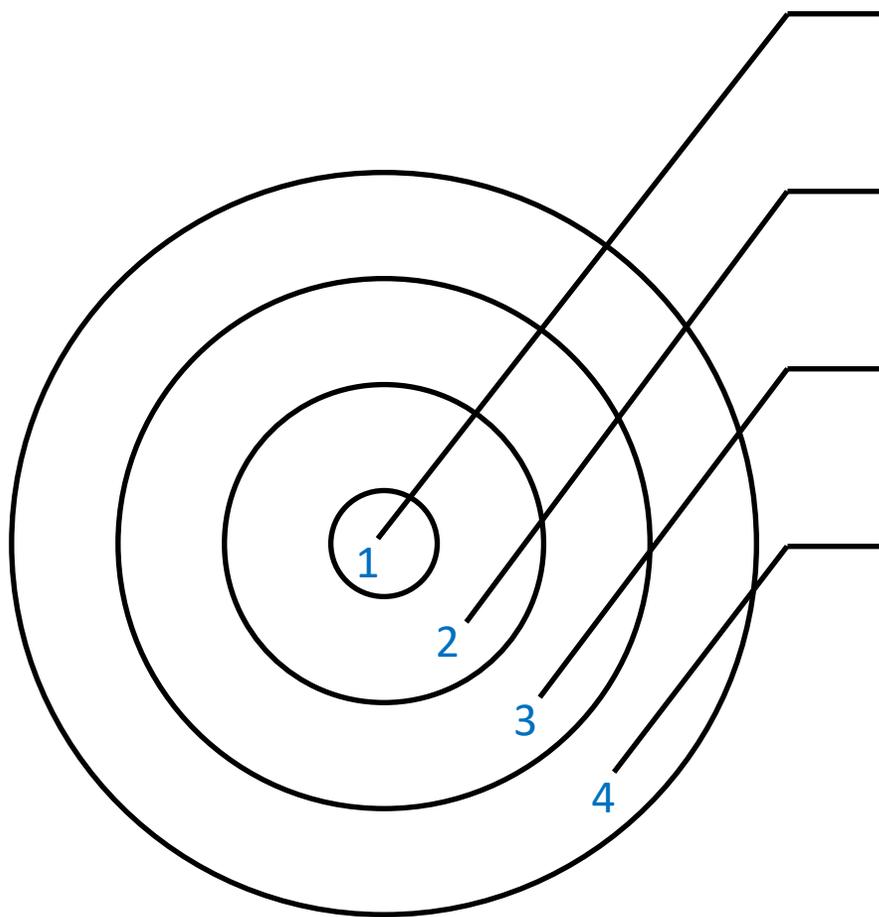
DiverSIta: obiettivi e milestone



1. **Linee guida** per una risorsa modulare, incrementale e replicabile per lo studio della lingua parlata
2. **Nuovi moduli** del corpus KIParla: KIPasti, ParlaBO, Stra-ParlaBO, Stra-ParlaTO
3. **Oral Compass**: una bussola dell'oralità. Materiali divulgativi basati sui testi del progetto e sviluppati insieme alle istituzioni locali (enti, associazioni, scuole, ecc.)
4. KIParla: **new release**



DiverSIta: obiettivi e milestone



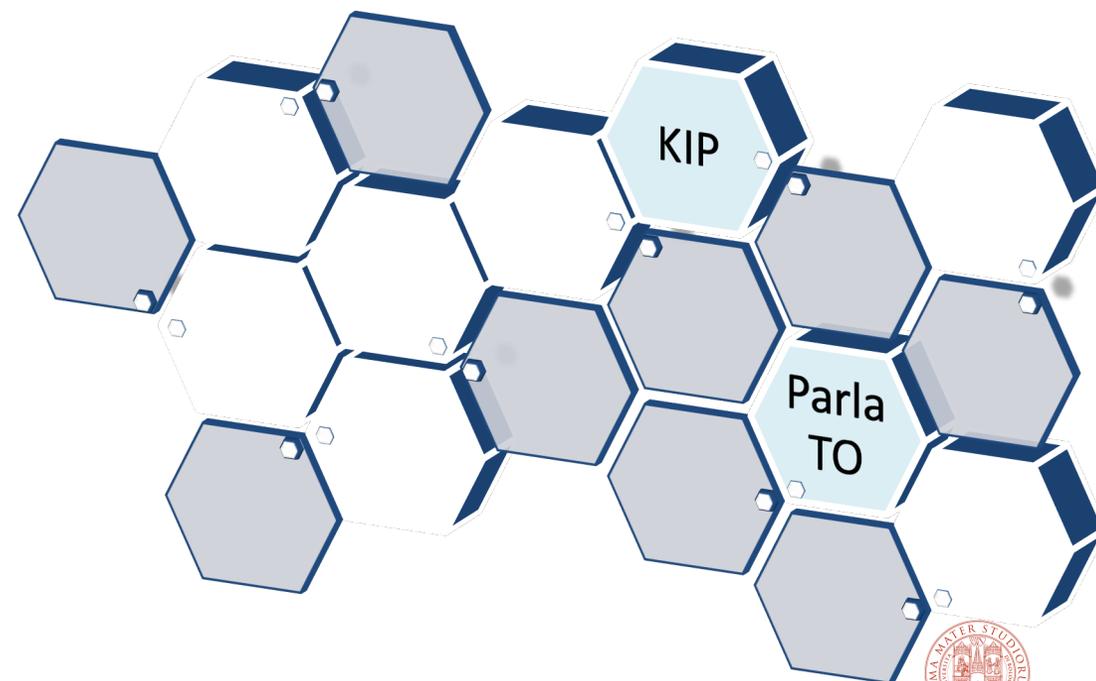
1. Linee guida

2. Nuovi moduli del corpus KIParla

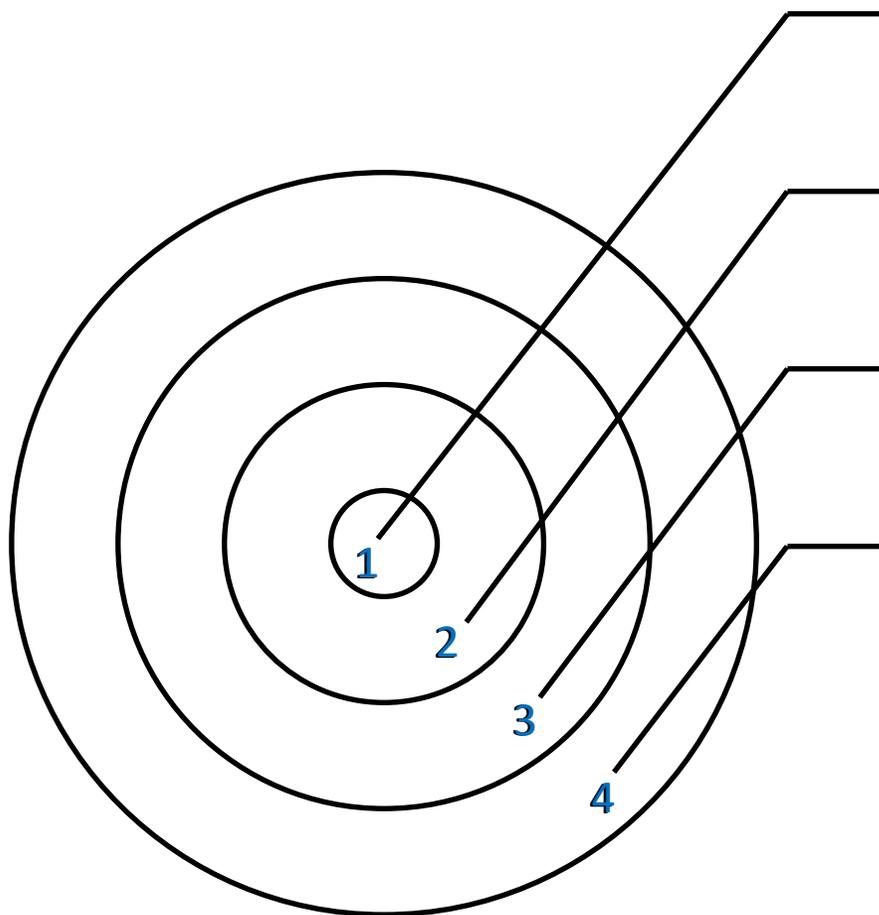
3. Oral Compass

4. KIParla

Già
pubblicati



DiverSIta: obiettivi e milestone



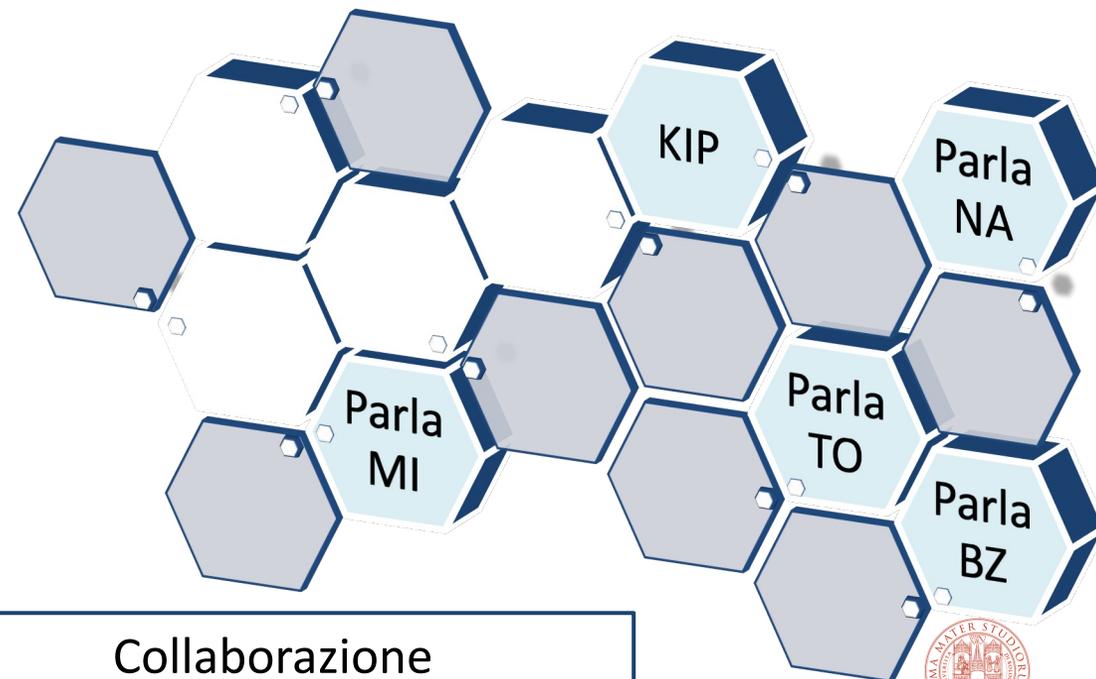
1. Linee guida

2. Nuovi moduli del corpus KIParla

3. Oral Compass

4. KIParla

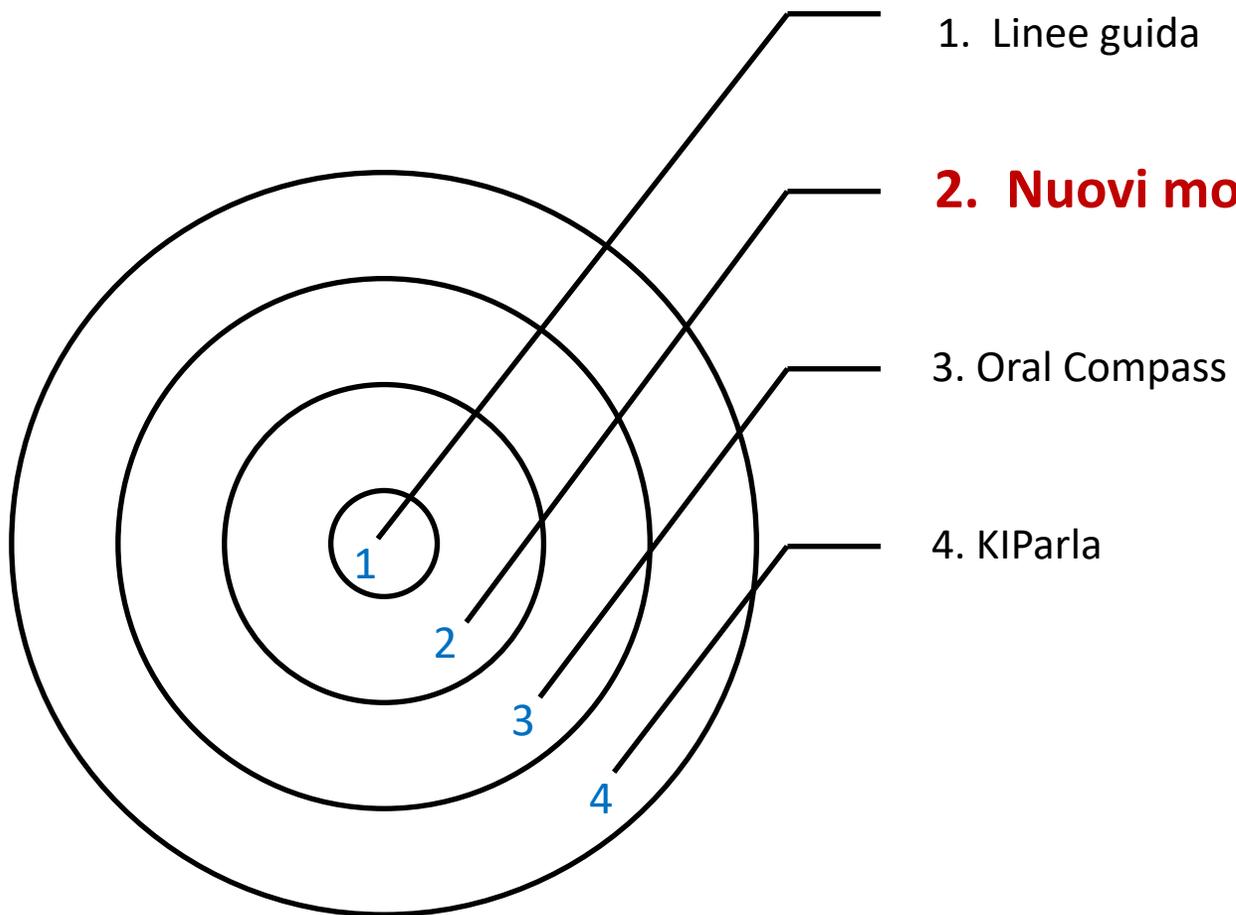
Collaborazioni
con BZ e NA
soon online



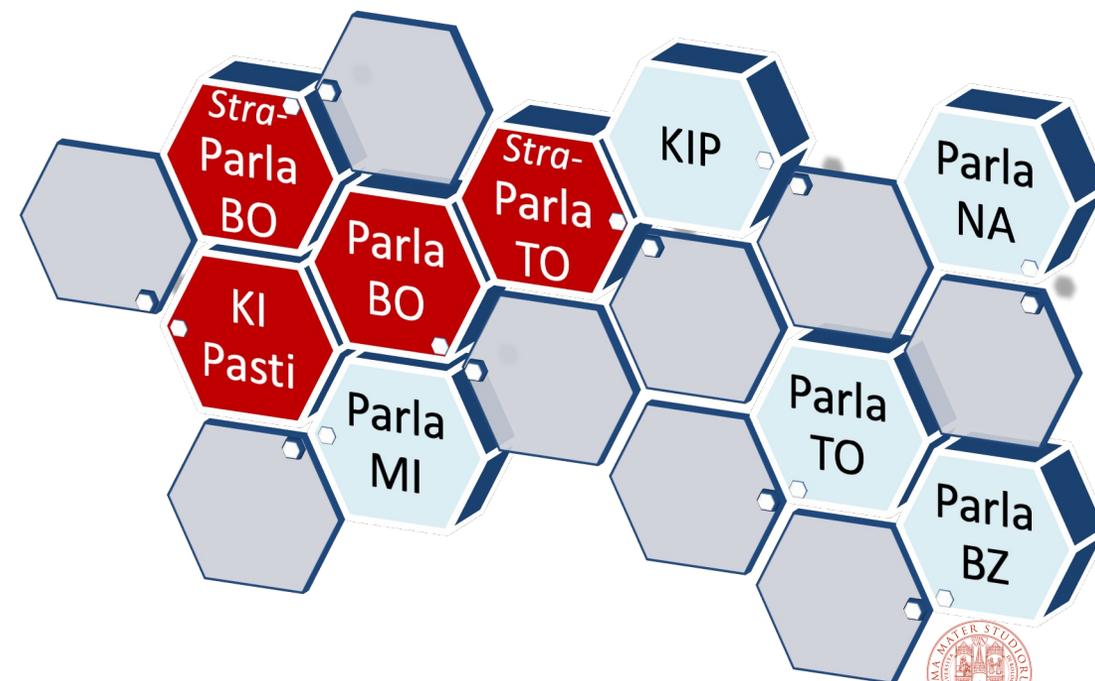
Collaborazione
con Milano Bicocca - *ongoing*



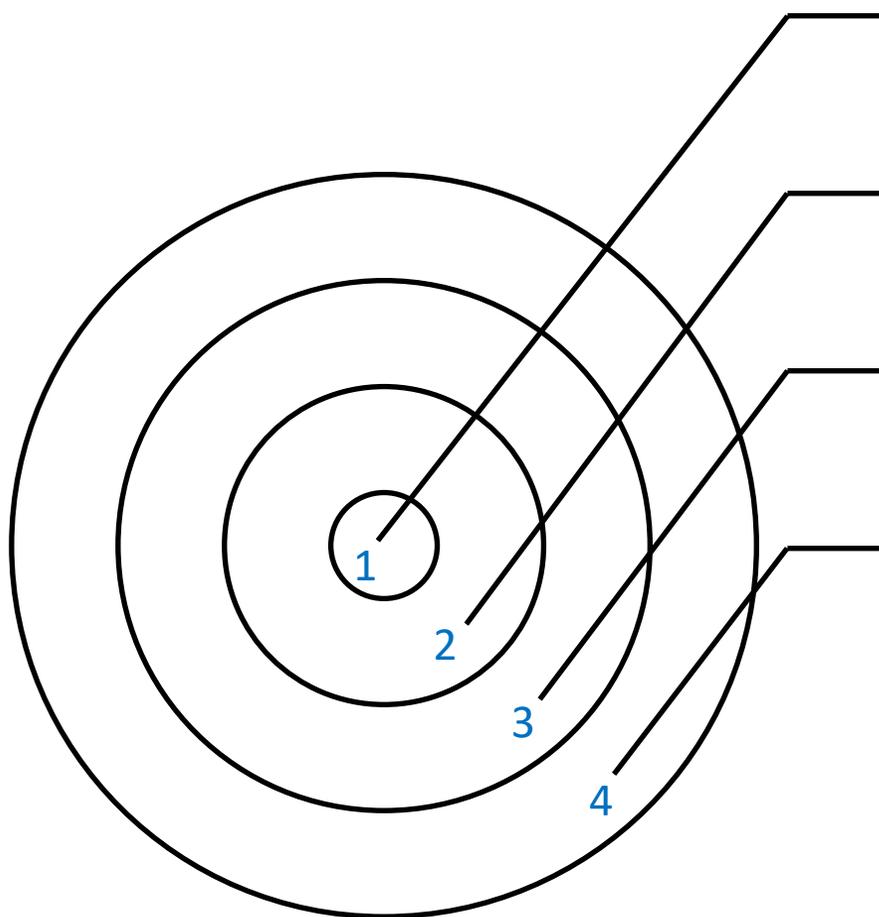
DiverSIta: obiettivi e milestone



Obiettivi del PRIN
DiverSIta



DiverSIta: obiettivi e milestone



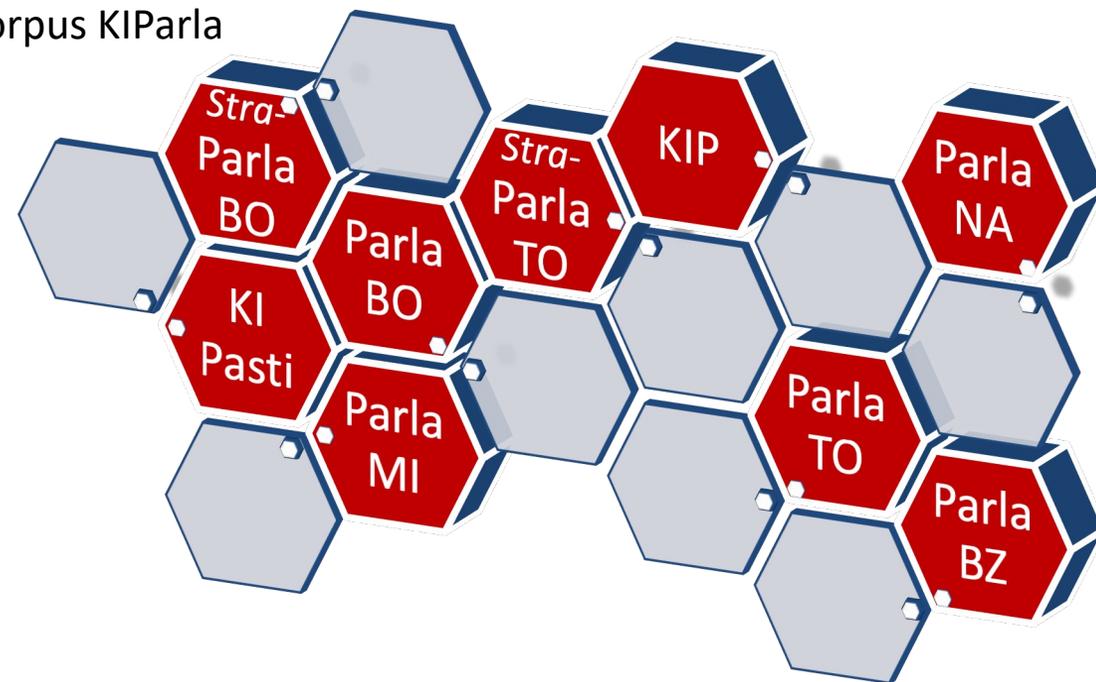
1. Linee guida

2. Nuovi moduli del corpus KIParla

3. Oral Compass

4. KIParla

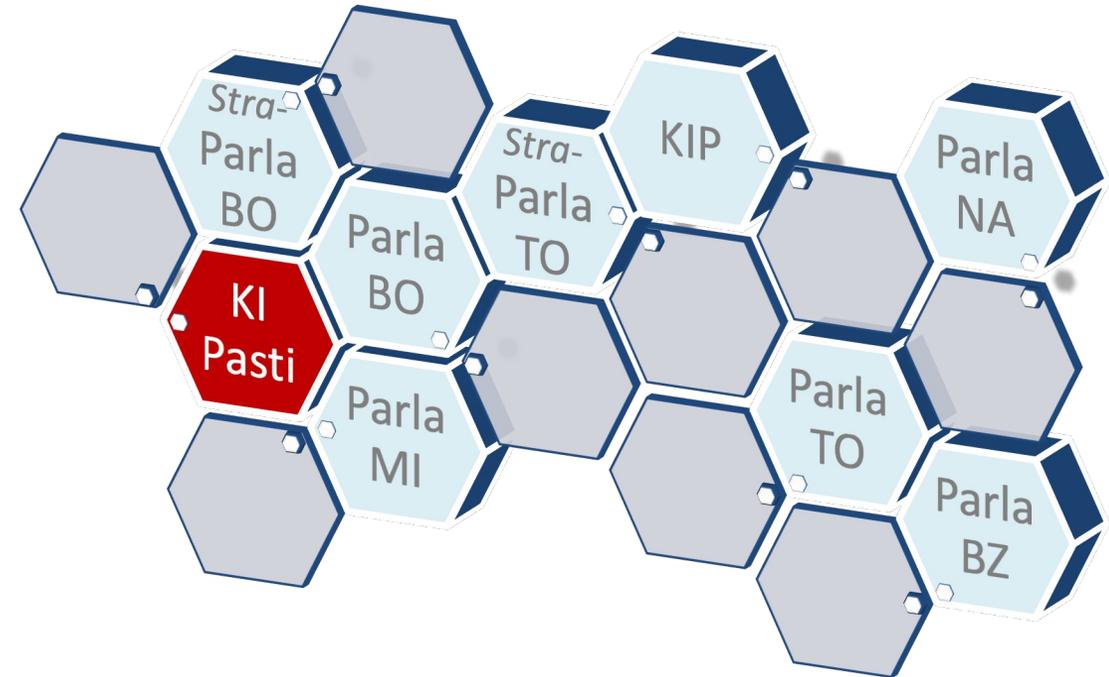
- Core set comune di metadati
- Trascrizione ortografica e Jefferson
- Allineamento con audio



- Moduli indipendenti e comparabili
- Search separata e congiunta
- ELAN > Python > NoSketchEngine



✓ KIPasti – conversazioni spontanee a tavola



✓ KIPasti – le conversazioni

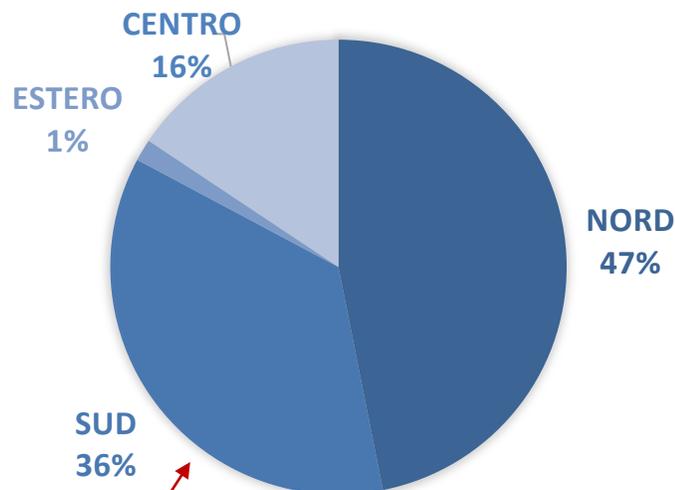
- ✓ 64 conversazioni a tavola
- ✓ 149 parlanti coinvolti
- ✓ 43:23 h di registrazione
- ✓ ca. 500.000 token

- Background condiviso
- Topic libero
- Registro familiare o amicale

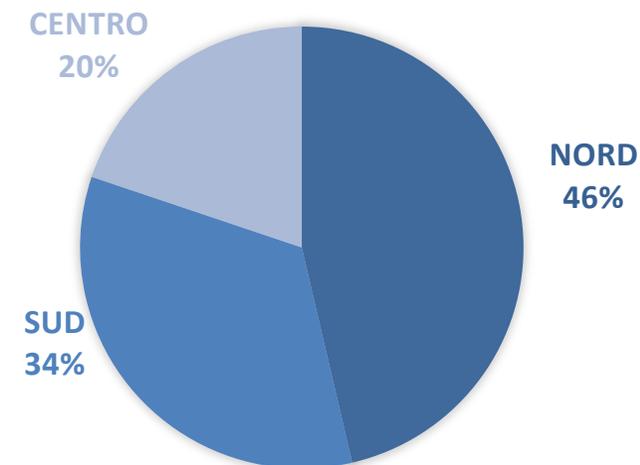
Bilanciamento: macro-area di registrazione

- Raccolta, trascrizione, anonimizzazione
- ✓ TERMINATE

RACCOLTA (KIPASTI)



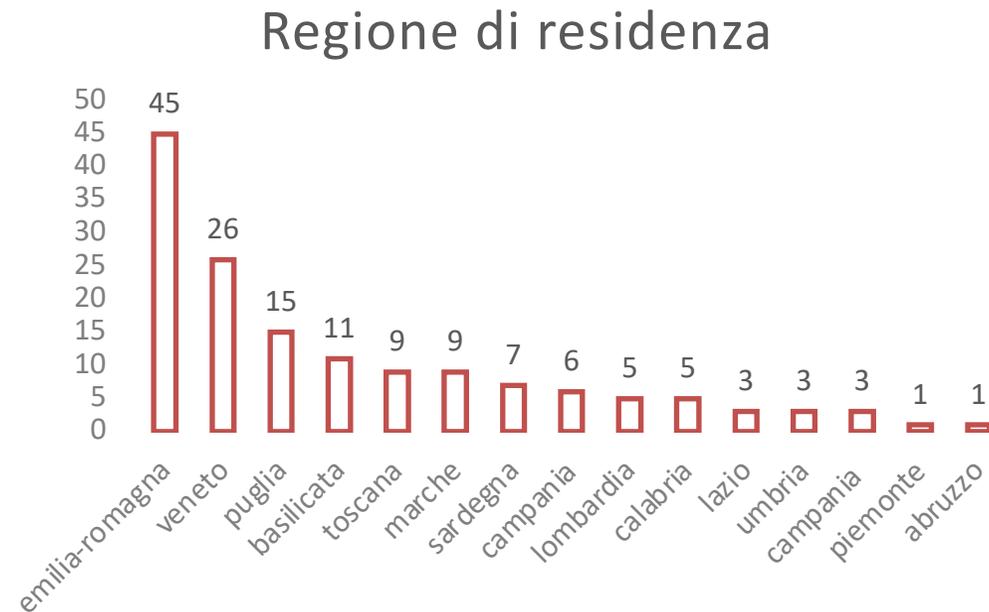
POPOLAZIONE (ISTAT)



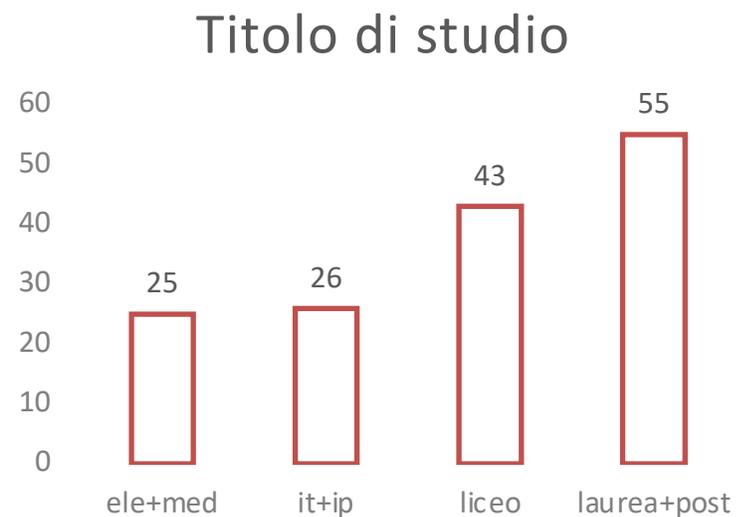
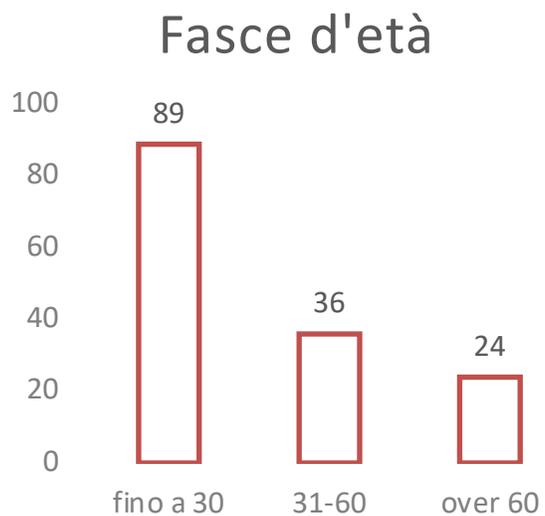
- 19 conversazioni registrate
- 8 conversazioni registrate
- 5 conversazioni registrate
- 3 conversazioni registrate
- 2 conversazioni registrate
- 1 conversazioni registrate



✓ KIPasti – i parlanti



Online
dal 15 aprile su
kiparla.it



KIPasti: alcuni esempi

KPC009 (Umbria)

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

PKP088 ((ride))

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]?

PKP090 [eh (.) ci cre:do].

PKP089 trenta: trenta euro.

PKP088 io pure:

PKP089 [mado:nna un colpo m'è pre:so] ,



Dati PKP089

Regione nascita: Umbria

Fascia di età: 21-25

Titolo di studio: diploma i.t.

Professione: area 6 (artigiani, operai specializzati e agricoltori)



KIPasti: alcuni esempi

KPN022 (Veneto)

PKP075 con le nuove normative (.) arriva il sindaco
(1.37)

e vi port- manda tutti a [casa] ((ride))

PKP076 ((ride)) [a consci]o viene il sindaco

PKP075 ((ride))

PKP076 ((ride))

PKP070 beh i pulotti possono arrivare lo ste[sso]

PKP076 [mh]mh quello sì

PKP075 i?

PKP070 pulotti
la polizia

PKP075 °ah°



Dati PKP070

Regione nascita: Veneto

Fascia di età: 21-25

Titolo di studio: diploma liceo

Professione: studente

Dati PKP075

Regione nascita: Veneto

Fascia di età: 16-20

Titolo di studio: licenza media

Professione: studentessa

Dati PKP076

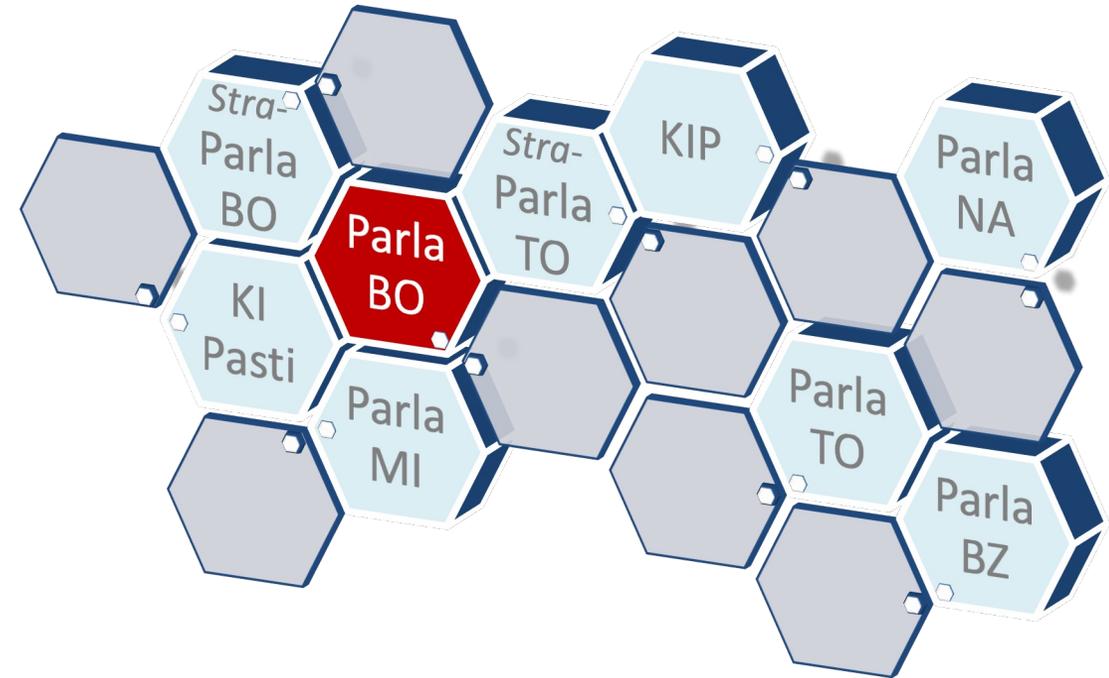
Regione nascita: Veneto

Fascia di età: 56-60

Titolo di studio: licenza media

Professione: area 6 (artigiani,
operai specializzati e agricoltori)

✓ ParlaBO – italiano parlato a Bologna



✓ ParlaBO – le interviste

- ✓ 86 interviste semi-strutturate
- ✓ 158 parlanti coinvolti
- ✓ 66:36 h di registrazione
- ✓ ca. 650.000 token

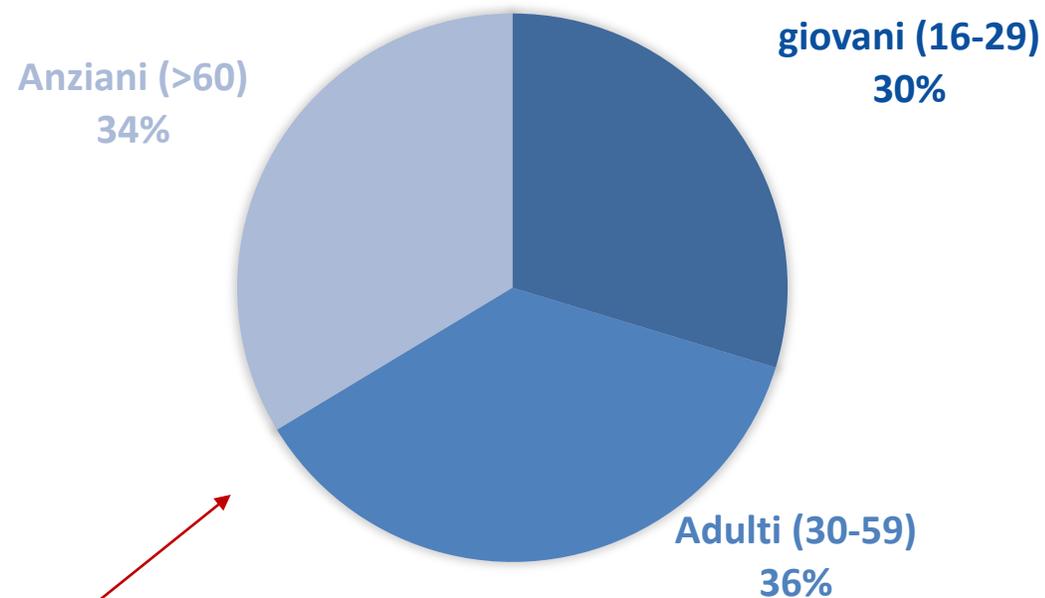
- Topic parzialmente guidato
- Registro per lo più informale
- 13 conversazioni includono l'uso del dialetto

Bilanciamento: fasce di età persone intervistate

Metadati disponibili (parlanti):

- ✓ regione di provenienza e residenza
- ✓ età
- ✓ titolo di studio
- ✓ professione

FASCE DI ETÀ



Argomenti intervista:

- ✓ rapporto con la città
- ✓ tradizioni, eventi, cibo
- ✓ lavoro
- ✓ tempo libero



✓ ParlaBO – i parlanti



- ✓ Raccolta dati terminata
- Attualmente in fase di trascrizione (70%)



Online entro giugno su kiparla.it



ParlaBO: alcuni esempi



PBB028 – alternanza con dialetto

racconto del terremoto in Emilia

BOI100 sono venuti i miei a chiamarmi:: (.) OU
#**livet bein** (.) eh c'è venuto il terremoto
#a:h i **dev andar a suner x matenna** ((ride))
((ride)) ho risposto così

PBB027 – tratti neostandard

racconto di un episodio dell'infanzia

BOI098 poi ci sono state diverse altre persone **dove gli** ha toccato la spalla
rossi un certo rossi
gli ha toccato la spalla ha preso il ponte di galliera



Dati PBB028

Regione nascita: Emilia-Romagna

Fascia di età: 31-35

Titolo di studio: diploma i.t.

Professione: area 6 (artigiani,
operai specializzati e agricoltori)

Dati PBB027

Regione nascita: Emilia-Romagna

Fascia di età: 31-35

Titolo di studio: diploma i.t.

Professione: area 6 (artigiani,
operai specializzati e agricoltori)



ParlaBO: alcuni esempi

PC007

racconto dell'arrivo degli alleati in Romagna, durante la Seconda Guerra

BOI019 eh sai l'inghi- l'inghilterra era una potenza anche::
nel nel in india dalle parti infatti lì c'erano degli indiani c'era
m- marocchini c'era tutta gente così
quando m::h
sono arrivati
oltre:: all'occupazione militare sono arrivati con le loro orchestre
quindi (.) e::hm nella nostra piazza (.) tutte le sere
c'era e:::hm:: un concerto
una musica nuova una musica che non avevamo mai sentito ed era la musica jazz

Dati BOI019

Regione nascita: Emilia-Romagna

Fascia di età: 91-95

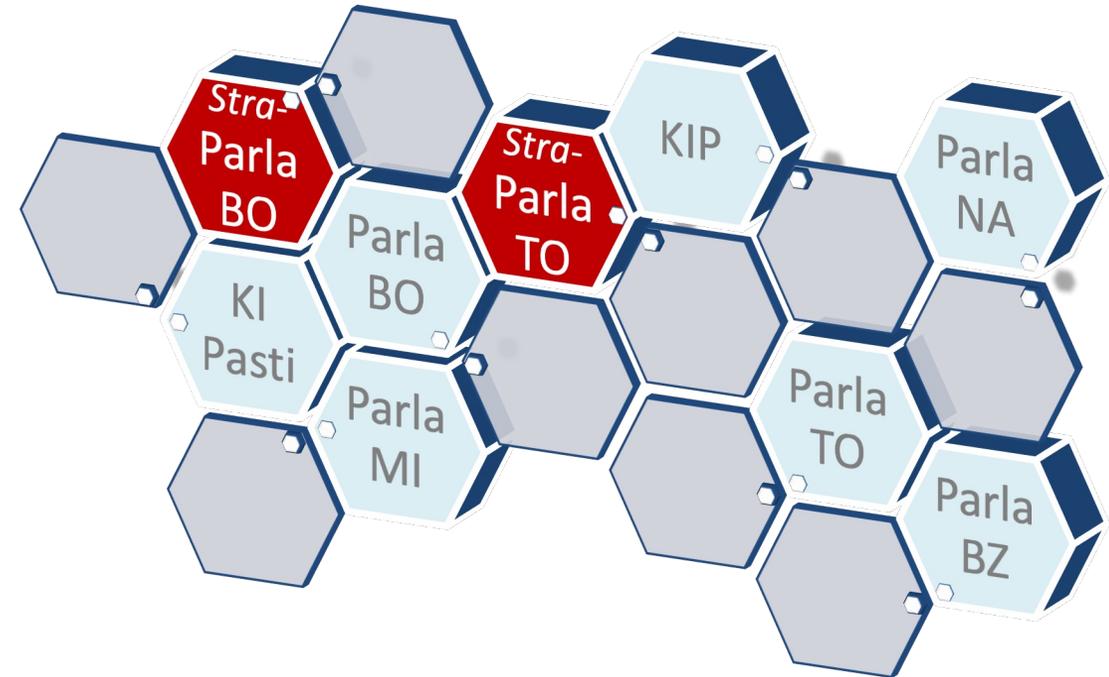
Titolo di studio: diploma liceo

Professione: area 2 (professioni intellettuali)



✓ ***Stra-ParlaBO* e *Stra-ParlaTO***

**italiano parlato a Bologna e Torino
da persone di origini non italiane
con un repertorio plurilingue**



La nuova raccolta dati: Stra-ParlaBO e Stra-ParlaTO

4 comunità per punto di raccolta *di cui 2 in comune (cinese e marocchina)*

BOLOGNA:

cinese, marocchina,
ucraina, bengalese

TORINO:

cinese, marocchina,
romena, peruviana

2 tipi di interazioni *comparabili con gli altri moduli del KIParla*

interviste
semi-
strutturate

conversazioni
spontanee
(pasti)

16 h di registrazione per comunità

8 h

Interviste

8 h

conversazione spontanea



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



Stra-ParlaBO e Stra-ParlaTO: l'intervista

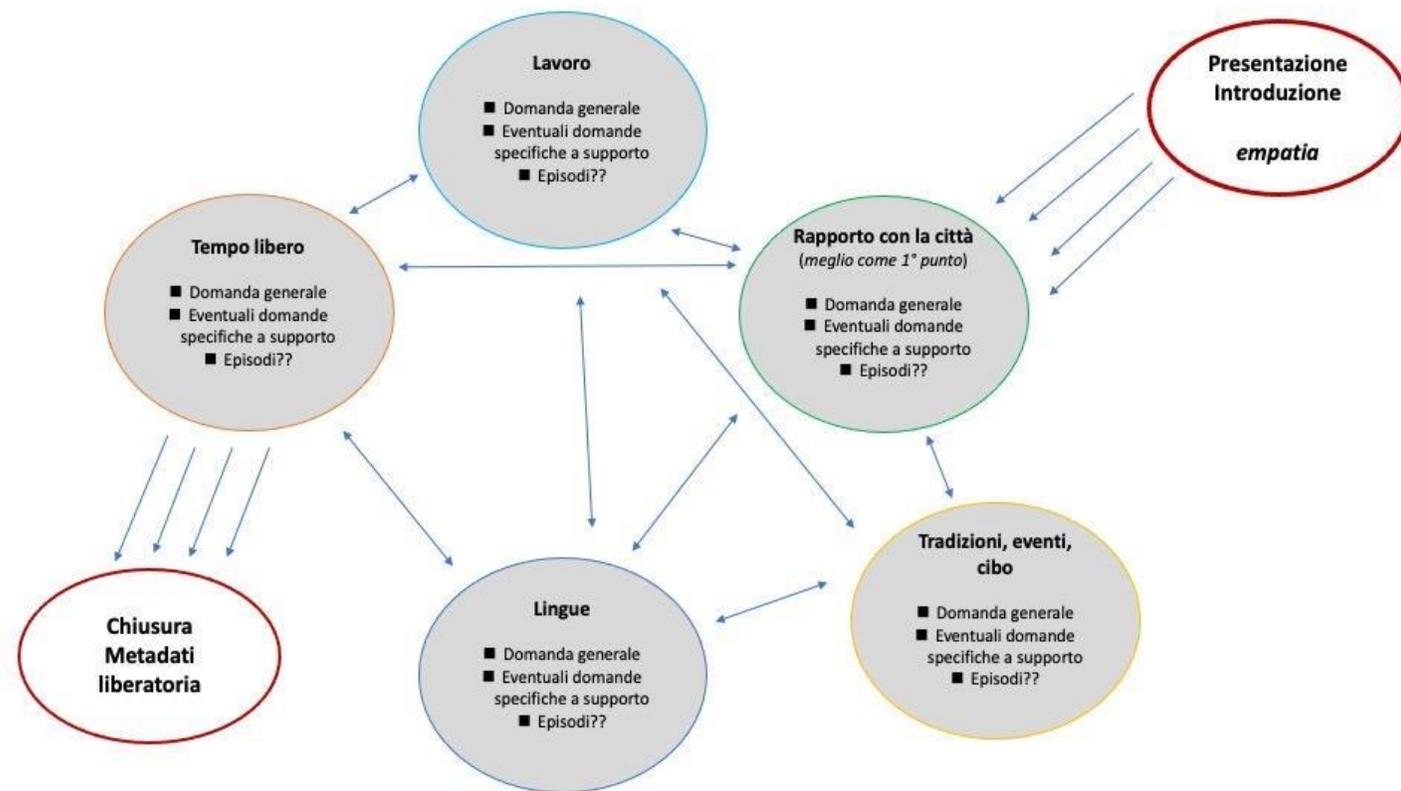
Blocchi tematici

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

Tema *lingue*:

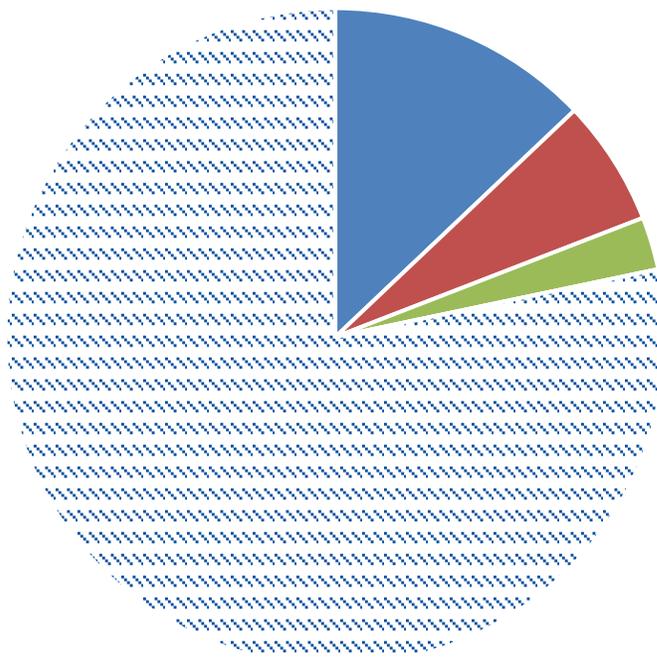
- Che lingue conosci?
- In quali contesti le parli?
- Che rapporto hai con l'italiano?
- Come e dove l'hai imparato?
- Che rapporto hai con la lingua d'origine?

Come affrontare l'intervista



Str-ParlaBO: dove siamo?

Stadio raccolta dati



■ C ■ U ■ A ■ B : da raccogliere

Comunità	Interviste	Pasti	Totale
Cinese	6:01:42	2:13:39	8:15:21
Ucraina	3:58:05	0	3:58:05
Marocchina	1:39:36	0	1:39:36
Bengalese	0	0	0
Totale	11:39:23	2:13:39	13:53:02



Stra-ParlaBO: un esempio



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

(0.95)

è è xx speciale eh

(2.77)

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

(0.88)

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]

(0.77)

PSB009 [da oc-]

PSB005 [vedere] sì

PSB009 [ve]dere da: occhi?

Dati PSB009

Comunità linguistica: cinese

Fascia d'età: 21-25

Titolo di studio: laurea

Professione: area 2 (professioni intellettuali)

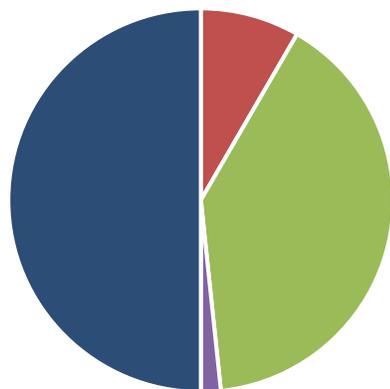
In Italia da: 4 mesi

Studentessa di scuola di italiano per stranieri



Stra-ParlaTO

Interviste



■ Cinese ■ Peruviana ■ Marocchina2 ■ Marocchina ■ Romena ■ Totale

Comunità	Interviste	Pasti	Totale
Cinese	0	0	0
Peruviana	02:05:00	0	02:05:00
Marocchina	00:25:00 + 10h circa	0	25:00 ± 10:00:00
Romena	0	0	0
Totale	02:30:00	0	02:30:00

Nel corso di una raccolta dati precedente sono già state raccolte 10 ore di intervista con persone nate sia in Marocco che in Italia. La loro integrazione nello Stra-ParlaTo è in corso di valutazione



Stra-ParlaTO: un esempio

TOR012 hai vissuto qualche, in qualche altro paese prima di arrivare in italia?
TOI102 no, e:hm
TOI102 # bəs fi-l-Mağrəb
traduzione solamente in Marocco
TOR012 sei andata a scuola prima di venire in italia?
TOI102 ehm, sì
TOI102 # l-ḥadd ə:: (.) rabaʕa: elementare
traduzione fino alla quarta elementare
TOR012 [con chi s-]
TOI102 #baʕdēn
traduzione dopo-
TOR012 # baʕdēn?
traduzione dopo?
TOI102 # baʕdēn get hīna. u- btedīt e:: la quinta elementare
traduzione poi sono arrivata qui. E ho iniziato la quinta elementare
TOR012 con chi sei venuta in italia?
TOI102 ehm, con mio fratello e mia mamma
TOI102 # ə: get hīna: ʕand ɓa:ɓa, w kān hīna::
traduzione e: sono venuta qui da mio padre, e sono stata qui
TOI102 # min: li- min is-sana:- (.) min il-duemila
traduzione da: a- dall'anno (.) dal duemila
TOI102 # fa-ḥna b-hīna nel duemilasette.
traduzione e noi siamo arrivati qui nel duemilasette



Dati PTC034

Comunità linguistica: marocco

Fascia d'età: 21-25

Titolo di studio: qualifica professionale

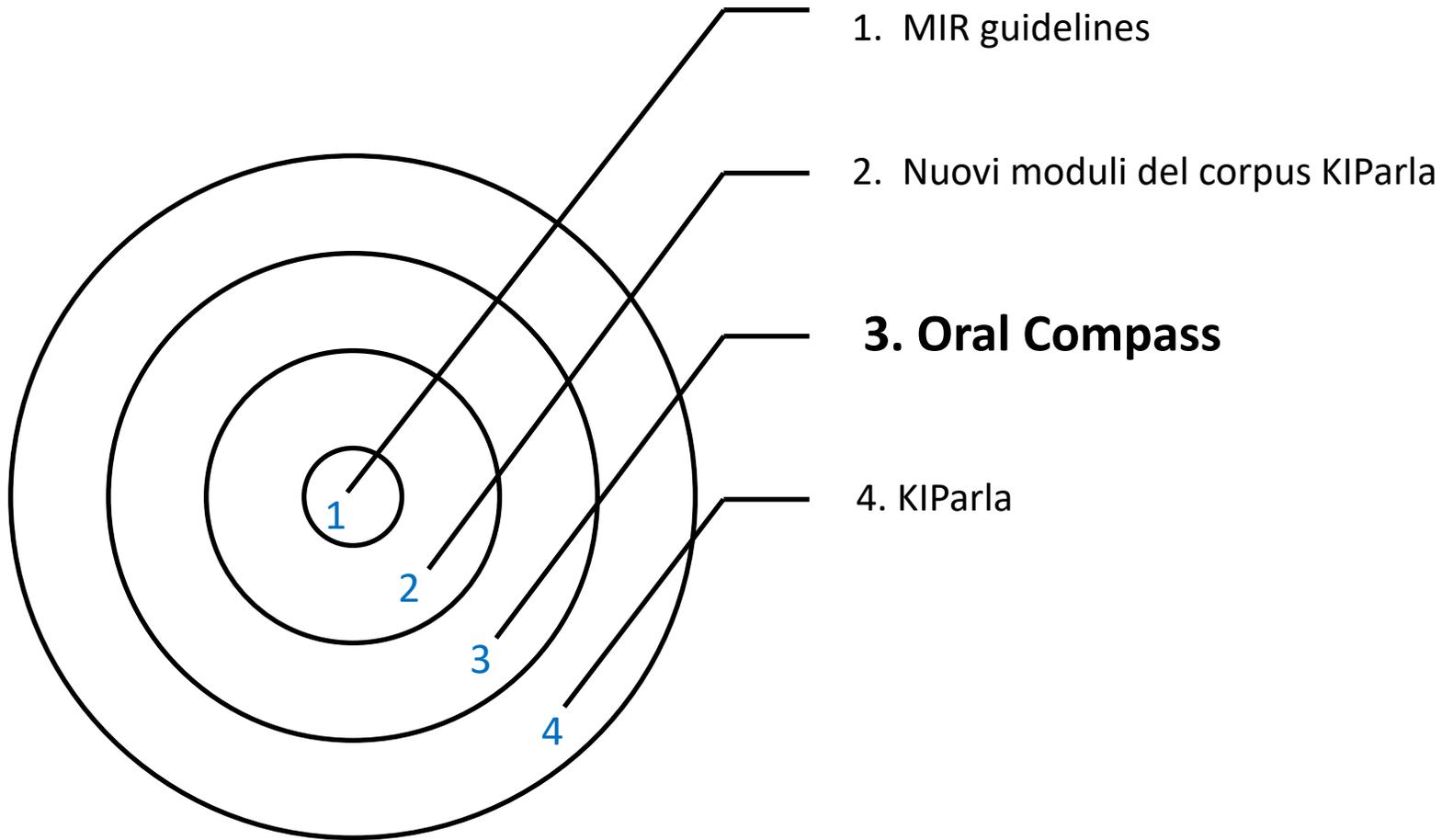
Professione: area 2 (professioni intellettuali)

In Italia da: 12 anni

Studentessa universitaria



DiverSIta: obiettivi e milestone

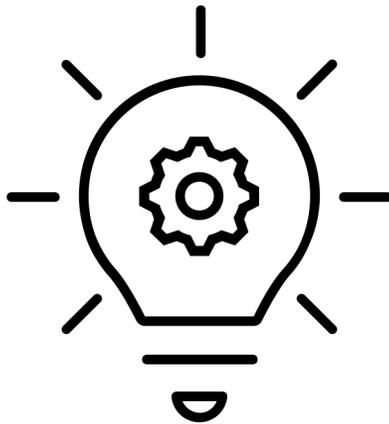


La bussola dell'oralità

- ✓ **Risorse e materiali basati sui testi del progetto e sviluppate insieme alle comunità (enti locali, associazioni, scuole, ecc.)**
- Prendere consapevolezza di come la lingua italiana viene usata per parlare della **diversità culturale**
 - Veicola **stereotipi**?
 - Fissa le persone in **categorie** rigide (es. 'gli stranieri', 'gli italiani', ecc.)
- Migliorare i servizi pubblici e del privato sociale



Alcune delle risorse della bussola



1. **Materiali per la formazione professionale**

- 'Io e la mia città multilingue'
- 'Noi vs loro: che terreno scivoloso!'
- 'With Italian... All I need is time'

2. **Linee guida per l'adattamento dei materiali in contesti specifici**

- Formazione a insegnanti
- Formazione in ambito sanitario
- Formazione nel settore dell'accoglienza
- Etc.

3. **Tutorial per l'uso autonomo del corpus a scopi di formazione**



Prossimi passi

Raccolta dati:

- proseguimento raccolta
- creazione collaborazioni con associazioni e altre realtà cittadine

Trascrizioni:

- adattamento del modello 'KIParla' ai nuovi dati
- esplorazione degli strumenti di AI per la trascrizione automatica (e.g. Whisper)

Annotazione del corpus:

- Lemmatizzazione
- POS tagging
- Parsing

**Ci vediamo a Torino
il 14 e 15 novembre!**





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