

**Una procedura informatica di accesso intelligente
a materiali in trascrizione fonetica:
l'esperienza dell'Atlante Lessicale Toscano**

**Luciano Agostiniani, Elisabetta Marinai,
Simonetta Montemagni, Matilde Paoli**

Centro di Documentazione del Lessico Toscano Moderno - Firenze

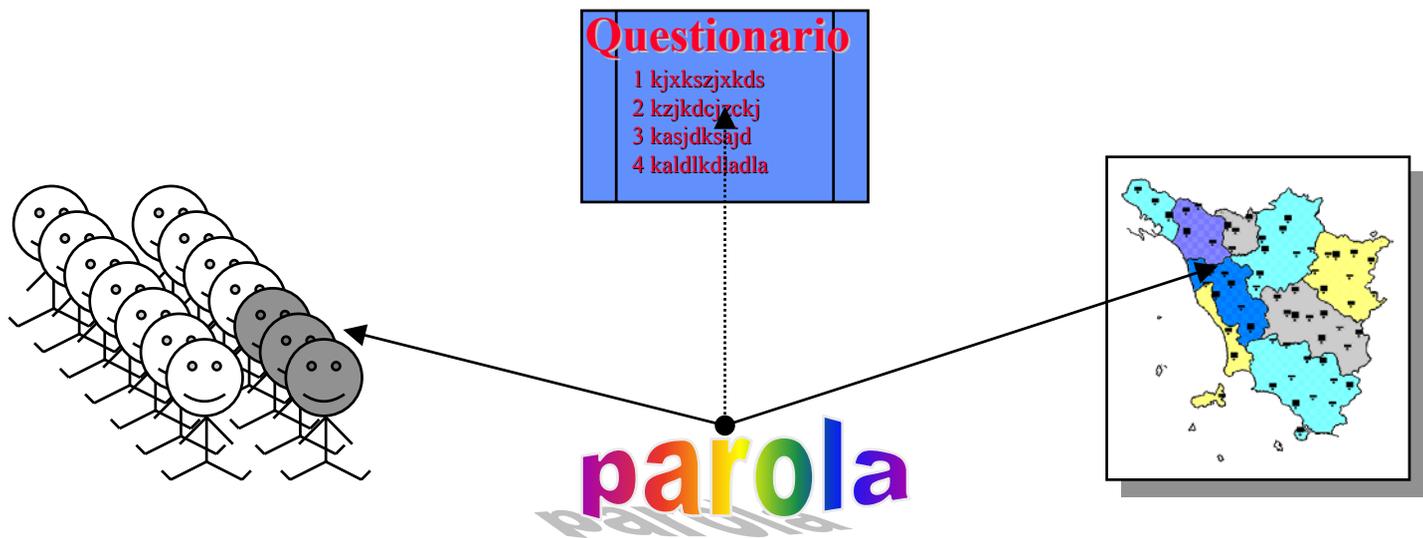
Istituto di Linguistica, Università degli Studi di Perugia

Istituto di Linguistica Computazionale, CNR - Pisa

L'Atlante Lessicale Toscano (ALT)

(redazione: Giacomelli [direttore], Agostiniani, Bellucci, Giannelli, Montemagni, Nesi, Paoli, Poggi Salani)

- L'ALT è un atlante linguistico in cui ogni dato lessicale è rapportato alle dimensioni **diatopica** e **diastratica**



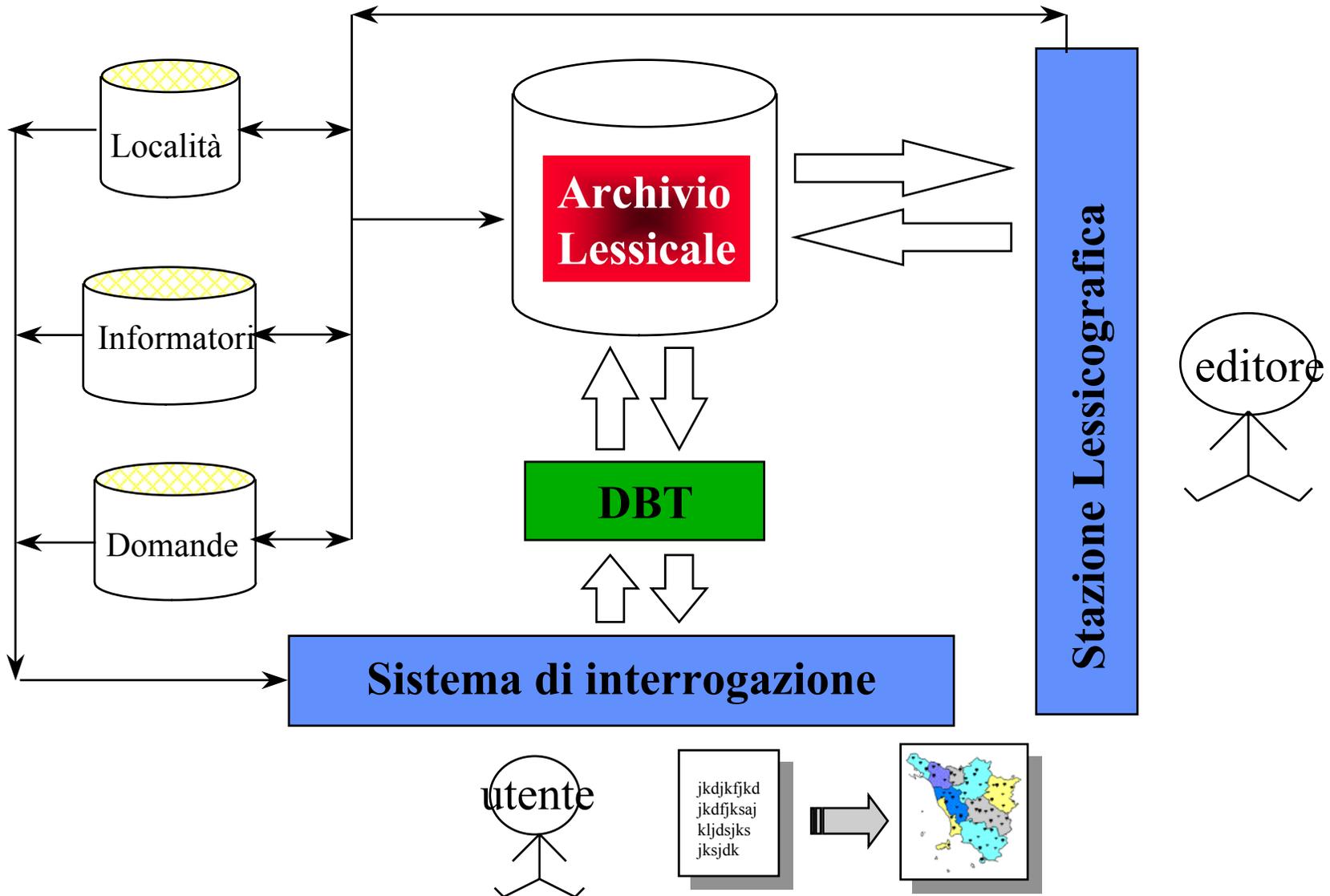
- La Banca Dati dell'ALT contiene i risultati di interviste svolte in **224** località della Toscana, con **2082** informatori sulla base di un questionario di **745** domande, per un totale di:
 - circa 350.000 schede che compattano le risposte alle domande del questionario
 - circa 30.000 schede contenenti materiali integrativi

L'ALT in versione elettronica: DBT-ALT

(Picchi, Biagini, Marinai, Merlitti)

- DBT-ALT è una versione specializzata del DBT (Eugenio Picchi, ILC-CNR), un sistema di database testuale per la memorizzazione, gestione ed interrogazione di grandi archivi di testi
- DBT-ALT gestisce una varia tipologia di dati linguistici strutturati, che contengono rappresentazioni sia in trascrizione fonetica sia in ortografia italiana
- Nella progettazione e sviluppo di DBT-ALT, si è inoltre dovuto far fronte alle specifiche esigenze della ricerca **geolinguistica** e **sociolinguistica** poste dall'ALT:
 - ☞ DBT-ALT gestisce un sistema integrato di archivi sussidiari contenenti informazioni riguardo a:
 - le località indagate
 - gli informatori intervistati
 - il questionario di raccolta
 - ☞ DBT-ALT offre anche la possibilità di proiettare su carta i risultati di una ricerca

DBT-ALT: architettura del sistema



DBT-ALT: chiavi di accesso ai materiali lessicali

- I dati dell'ALT sono oggi interrogabili secondo le canoniche chiavi di accesso ad un corpus di materiali dialettali raccolti sul campo tramite questionario:
 - ➡ la domanda di cui i materiali costituiscono risposta
 - ➡ la località in cui sono stati raccolti
- Inoltre, gli stessi dati possono essere **filtrati** sulla base delle caratteristiche socio-culturali dell'informatore che li ha attestati
- A queste modalità di accesso se ne affiancano altre che permettono esplorazioni personalizzate, a partire ad esempio da
 - ➡ una chiave semantica
 - ➡ una attestazione dialettale
- L'accesso a partire da attestazioni dialettali, che sono riportate in trascrizione fonetica, costituisce l'oggetto di questa presentazione

Tipologia dei materiali in trascrizione fonetica dell'ALT

```
{ Punto } 026  
{ TpInc } O  
  { Dom } 094  
{ Inf. A } 1  
{ Forma } <sékkatôjo>  
{ CGram } SO
```

```
{ Punto } 062  
{ TpInc } O  
  { Dom } 001b  
{ TpNot } F  
  { Inf. A } 1  
{ Testo } <la fálče la fá le púnte in vètta e i  
          kkôrho in fôndo, la jáče>
```

```
{ Punto } 026  
{ TpInc } O  
  { Dom } 094  
{ Inf. A } 2,3,4,5,6,7,8,9,B  
{ Forma } <sékkadôjo>  
{ CGram } SO
```

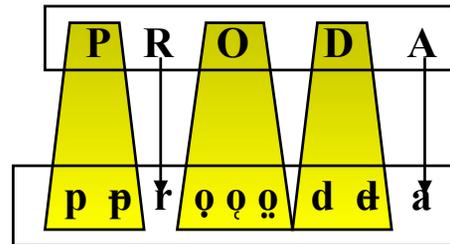
{ Descr } Intero edificio adibito alla funzione di seccatoio, costituito da un unico, non molto ampio, locale, di solito quadrato, articolato su due piani: al centro di quello inferiore ci sono i <lastrôni> su cui si accende il fuoco, mentre il 'soffitto' di questo stesso piano ha una struttura portante di <trávi> di legno su cui poggia un graticcio di <astôni>, sul quale si dispongono le castagne e al quale si accede da un <finestrône>.

Recupero di materiali lessicali in trascrizione fonetica: generalità

- Le difficoltà di recupero di materiali in trascrizione fonetica sono generalmente risolte tramite procedure di tipizzazione che associano ad ogni forma trascritta un corrispondente tipo “astratto” che viene assunto ad oggetto della ricerca
- Ma la tipizzazione di grosse quantità di dati, come nel caso del corpus dei materiali ALT (~400.000 schede contenenti attestazioni di varia natura), pone problemi pratici ed anche teorici
- I criteri di tipizzazione possono variare a seconda:
 - della tipologia dei materiali da tipizzare
 - dell’obiettivo della ricerca
- L’impossibilità di tipizzare l’intero corpus dei materiali ALT prima della sua pubblicazione ci ha portato a dover affrontare il problema del recupero di dati trascritti foneticamente senza una preventiva elaborazione manuale dei materiali

Recupero automatico di materiali lessicali in trascrizione fonetica senza tipizzazione preventiva

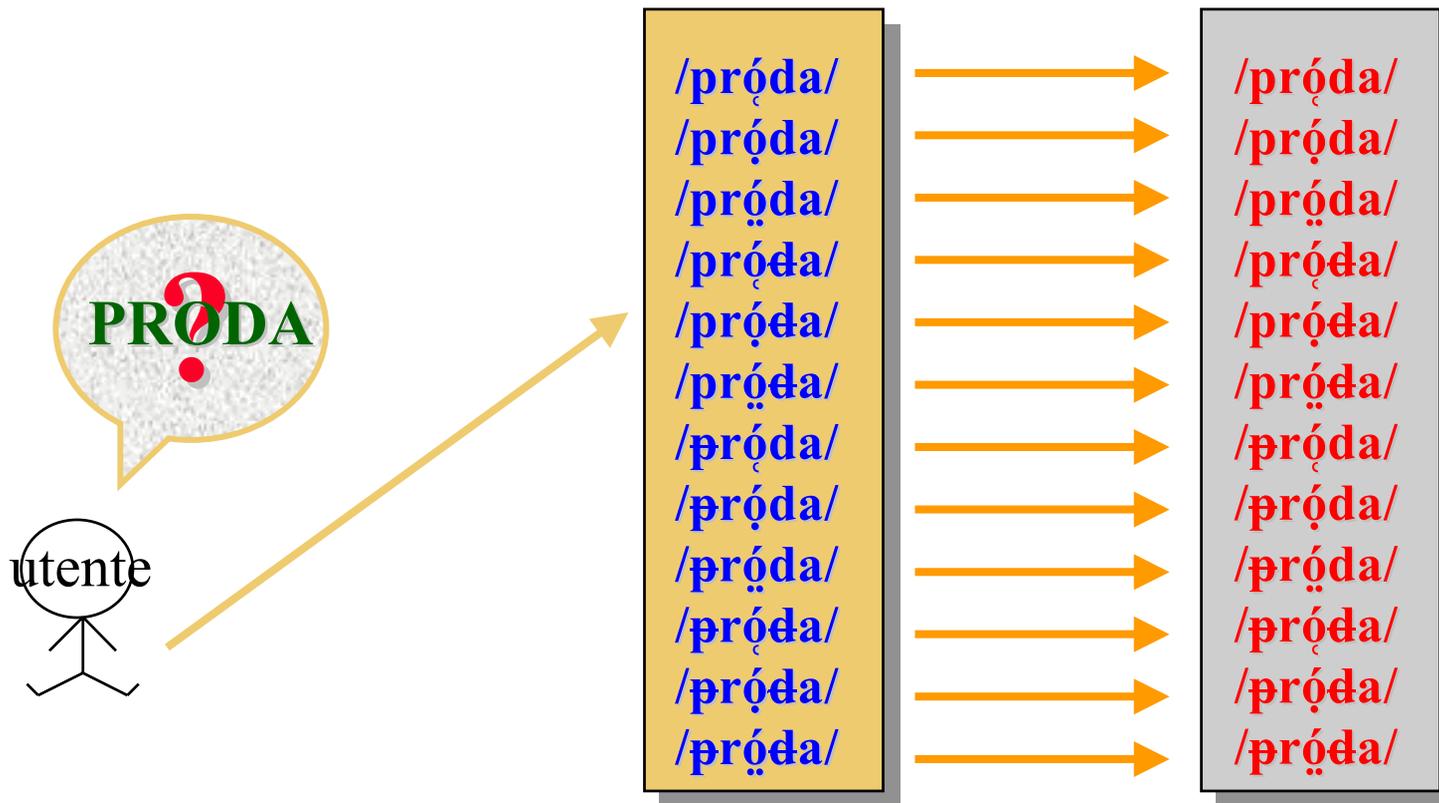
- Le procedure canoniche di accesso automatico ai dati di un archivio testuale presuppongono in partenza una **corrispondenza forte** (cioè **esatta**) tra l'oggetto della ricerca e quanto recuperato
- Nel caso di materiali in trascrizione fonetica, si dovrebbe conoscere in anticipo la realizzazione fonetica esatta delle parole oggetto della ricerca



- Una possibile alternativa: le “fuzzy matching techniques”, cioè tecniche che non impongono la restrizione di una corrispondenza forte tra l'oggetto della ricerca e l'informazione recuperata (**corrispondenza debole**, cioè **approssimata**)
- Queste tecniche, se da un lato accrescono la quantità di informazioni recuperate, dall'altro introducono molto rumore tra i risultati della ricerca
- Esempio: PRODA chiesto come PR??A porta al recupero di forme come: prima, preda, prova, prosa, etc.

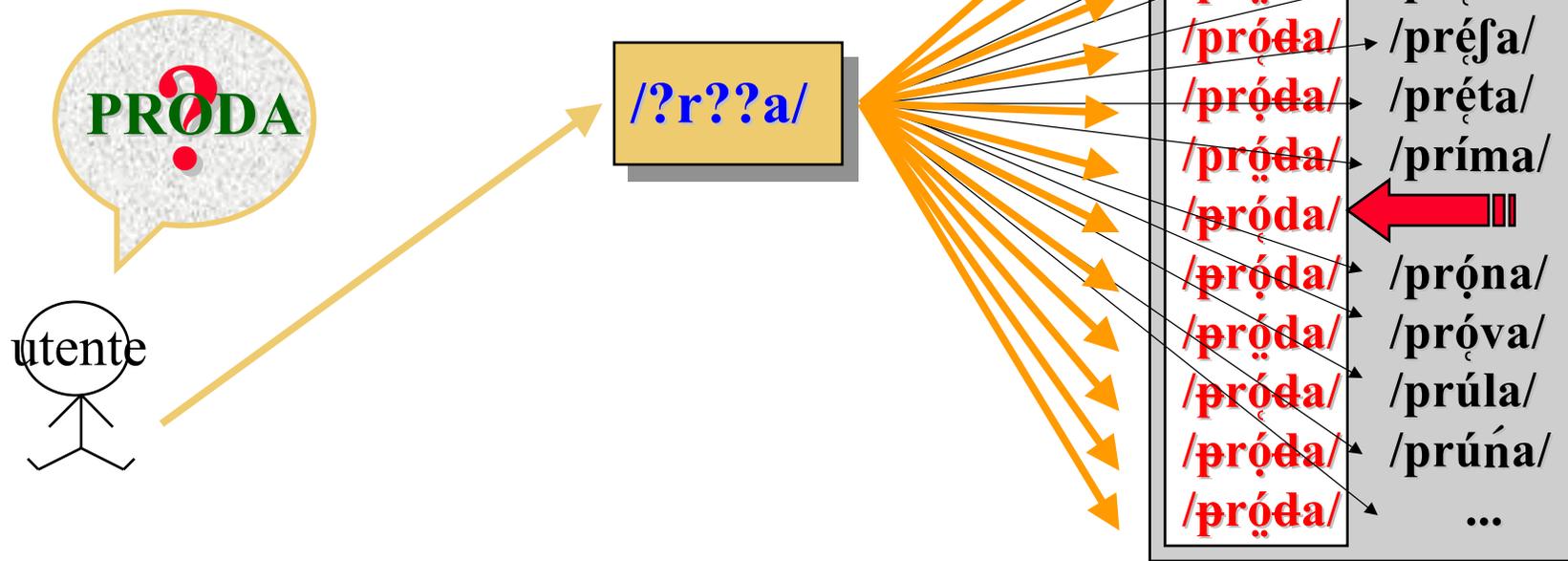
Recupero automatico di materiali lessicali in trascrizione fonetica senza tipizzazione preventiva (1)

- Le procedure canoniche di accesso automatico ai dati di un archivio testuale presuppongono in partenza una **corrispondenza forte (cioè esatta)** tra l'oggetto della ricerca e quanto recuperato
- Nel caso di materiali in trascrizione fonetica, si dovrebbe conoscere in anticipo la realizzazione fonetica esatta delle parole oggetto della ricerca



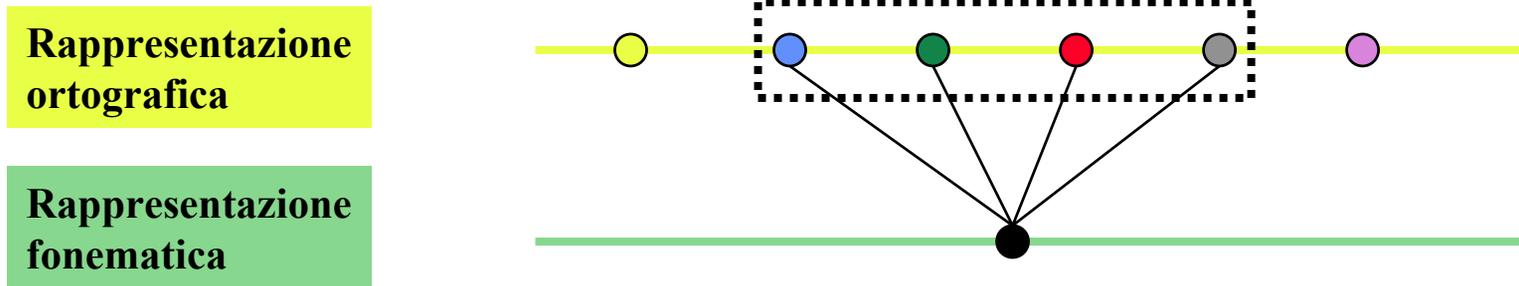
Recupero automatico di materiali lessicali in trascrizione fonetica senza tipizzazione preventiva (2)

- Una possibile alternativa è costituita dalle “fuzzy matching techniques”, tecniche che non impongono la restrizione di una corrispondenza forte tra l’oggetto della ricerca e l’informazione recuperata (**corrispondenza debole**, cioè **approssimata**)
- Queste tecniche, se da un lato accrescono la quantità di informazioni recuperate, dall’altro introducono molto rumore tra i risultati della ricerca



Tecniche esistenti di corrispondenza debole basate su similarità fonetica

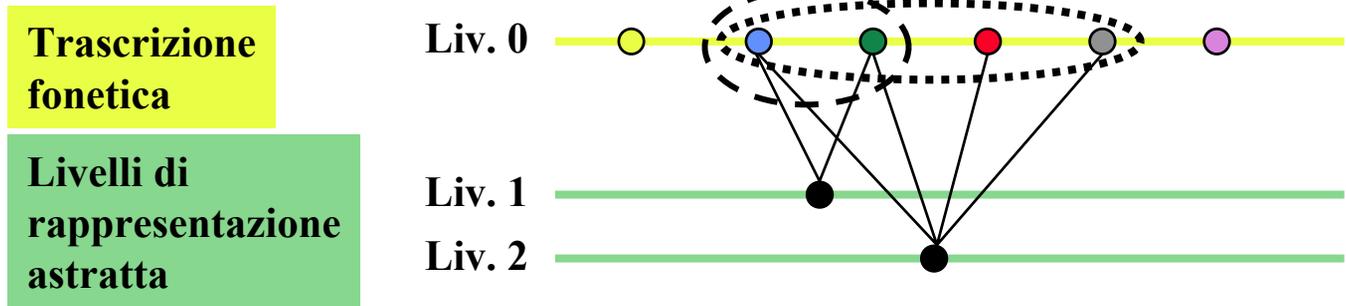
- I sistemi di correzione ortografica e i sistemi per l'accesso a grosse banche dati di nomi e indirizzi, nel caso di mancata corrispondenza forte, cercano forme vicine a quella ricercata facendo uso di tecniche di corrispondenza debole basata su similarità fonetica



- Esempi: {Smith, Smit, Smythe, Smeeth}, {Lenci, Lenzi}, {Balsini, Balzini}
- Questi sistemi effettuano un riallineamento della rappresentazione ortografica e di quella fonemica tramite la creazione di classi di equivalenza di grafemi o sequenze di grafemi (in contesto)
- In questo modo, il rumore introdotto dalle “fuzzy matching techniques” viene ridotto in modo significativo migliorando qualità e precisione del risultato della ricerca

Una possibile soluzione per il recupero di materiali lessicali in trascrizione fonetica non tipizzati

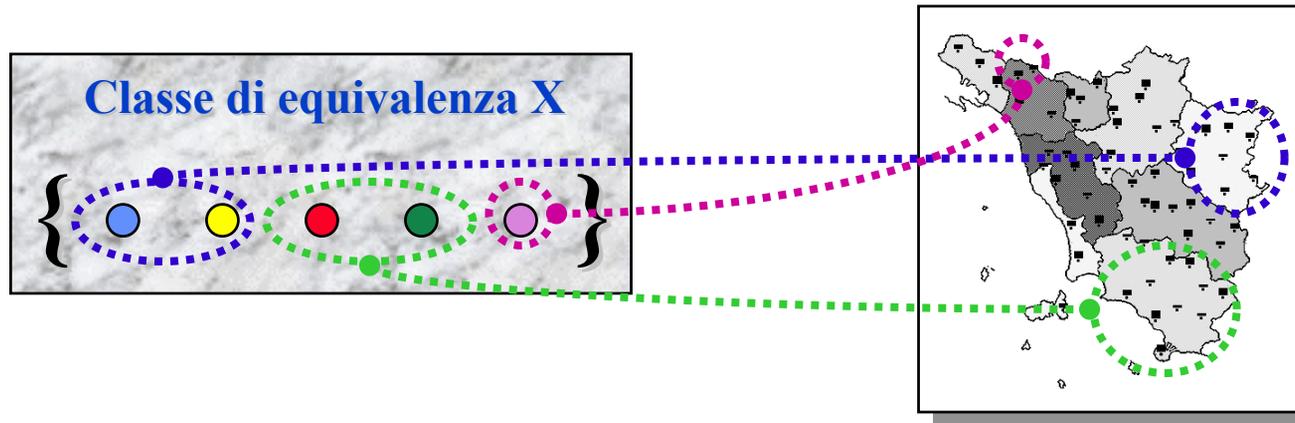
- Nello sviluppo di DBT-ALT, è stata messa a punto una procedura di accesso intelligente ai dati in trascrizione fonetica dove la corrispondenza debole tra la forma oggetto della ricerca e quelle recuperate è calcolata sulla base della conoscenza dei dialetti toscani



- Tale procedura offre all'utente la possibilità di selezionare livelli differenziati per l'accesso ai dati:
 - ☞ un primo livello (0) che richiede una **corrispondenza forte** tra l'oggetto della ricerca e il dato recuperato
 - ☞ una serie di livelli che si basano su **corrispondenze deboli**, permettendo così all'utente di astrarre da tratti specifici della realizzazione fonetica:
 - livello 1, che astrae da segni diacritici quali apertura di vocale, spirantizzazione di occlusiva etc.
 - livello 2, che si basa su relazioni complesse di equivalenza tra fonie (semplici e composte)
 - livello 3, che astrae da variazioni di tipo morfologico
 - livello 4, basato sulla struttura sillabica (scheletri consonantici/vocalici)

Dietro alla procedura di accesso intelligente a dati in trascrizione fonetica dell'ALT

- Il sistema di trascrizione fonetica adottato nell'ALT si rifà al sistema di trascrizione stabilito per la Carta dei Dialetti Italiani (CDI) in versione specializzata per la codifica di materiali relativi ai dialetti toscani
- Le corrispondenze deboli calcolate per l'accesso di livello 1 e 2 sono basate sulla definizione di **classi di equivalenza di fonie**, foni singoli o sequenze di foni
- Le classi di equivalenza definite raccolgono alternanze di fonie ricorrenti nella varietà dei dialetti toscani (ad esclusione delle aree marginali), sia all'interno dello stesso sistema dialettale sia tra sistemi diversi



Classi di equivalenza di foni di livello 1

CLASSE di EQUIVALENZA	BASE	CLASSE di EQUIVALENZA	BASE
a, ä ($\pm \sim, \pm'$)	a	k, k'	k
b, ɸ	b	l, ɭ, ɮ	l
č, č̣	č	n, ɲ	n
d, ɖ	d	o, ɔ, ɔ̣, ɔ̤, ɔ̥, ɔ̦, ö ($\pm \sim, \pm'$)	o
e, ɛ, ɛ̣, ɛ̤, ɛ̥, ɛ̦, ë ($\pm \sim, \pm'$)	e	p, ɸ	p
ɛ̇		š, ś	š
ǰ, ǰ̣	ǰ	ǰ, ǰ̣	ǰ
g, ɣ	g	t, ɖ	t
i ($\pm \sim, \pm'$)	i	u, ü ($\pm \sim, \pm'$)	u
ĩ		ũ	

Queste classi di equivalenza sono definite attraverso un sistema di codifica della trascrizione fonetica che si basa su **rappresentazioni composizionali** dove ogni simbolo dell'alfabeto fonetico è codificato mediante una base che può essere ulteriormente specificata per mezzo di diacritici (ad esempio, ɛ̇ = e1)

Classi di equivalenza di foni di livello 1 e recupero di dati lessicali

Tratti fonici neutralizzati:

Esempi:

-
- | | |
|---|--|
| ● grado di apertura della vocale | /pr ^o da/, /pr ^o da/, /pr ^o da/ |
| ● spirantizzazione di occlusiva | /ab ^e to/, /ab ^e to/ |
| ● perdita di occlusione nelle affricate | /a ba ^č io/, /a ba ^č io/ |
| ● nasalizzazione di vocale | /b ^o mba/, /b ^o mba/ |
| ● turbamento di vocale | /l ^u na/, /l ^u na/ |
| ● consonantizzazione di vocale | /vi ⁱ otto/, /vi ⁱ otto/
/č ^e duo/, /č ^e duo/ |
| ● carattere velare di /l/ e /n/ | /alta ^l ena/, /alta ^l ena/ |
| ● presenza dell'accento | /kr ^o nnolo/, /kr ^o nnolo/ |
-

Classi di equivalenza di fonie di livello 2 e recupero di dati lessicali

CLASSE di EQUIVALENZA	ESEMPI
č, čč, č, š (sse #_V // V_V)	/a bačío/, /a baččío/, /a bačío/, /a bašío/
ǰ, ǰ, ǰǰ, ǰ (sse #_V // V_V)	/ǰákka/, /ǰǰákka/, /la ǰákka/, /ǰákka/
k, k', h	/tornáre di kása/, /- di k'ása/, /- di hása/
kǰ, k'ǰ, kkǰ, tǰ, tǰ, ttǰ, č, čč, t', t't'	/mákkǰa/, /máča/, /máčča/, /mát't'a/ /grańńolískǰio/, /grańńolístǰio/, /grańńolísčio/
ǰi, ǰi, ǰǰi, dǰ, dǰ, dǰǰ, ǰ', ǰ'ǰ', j, ǰǰ	/ǰiaččǰaia/, /dǰiaččǰaia/, /ǰ'iaččǰaia/, /d'aččǰaia/
l, r (sse _C[≠r]), ǰ (sse _C _i C _i), l' (sse _C)	/dólko/, /dórko/, /dól'ko/, /dól'kko/
ll, d, dǰ	/ballótti/, /bađótti/
l', l'l', li, lli, ǰ, ǰǰ, d', d'd'	/číl'l'o/, /číl'li'o/, /číl'ǰǰ'o/
ń, ńń, ni, nni	/pańńone/, /pań'ni'one/
r, rr (sse V_V)	/karraréčča/, /kararéčča/
s, š, ś, š (sse _C), ž (sse _C), z (sse l_ // r_ // n_)	/esóso/, /ešóšo/ /bǰastéma/, /bǰástéma/, /bǰaštéma/ /ánsimo/, /ánzimo/
š, šš	/kášša/, /káša/
z, zz, 3, 33	/zzázžera/, /zážžera/, /33ázžera/, /33ázžera/

Classi di equivalenza di livello 1 e 2 a confronto

- Classi di equivalenza di **livello 1**:
 - ☞ operano a livello del **singolo fono**
 - ☞ l'appartenenza di un dato fono ad una classe specifica **non è condizionata dal contesto** in cui occorre
 - ☞ le classi definite sono **disgiunte**
- Classi di equivalenza di **livello 2**:
 - ☞ operano sia a livello del **singolo fono** sia di **sequenze di foni**
 - ☞ l'appartenenza ad una classe specifica può essere **contestualmente condizionata**
 - ☞ le classi definite possono presentare **intersezioni**
- In entrambi i casi, l'alternanza è ricercata sia a livello di singola parola sia in contesto di frase (si astrae così da fenomeni quali il raddoppiamento fonosintattico)
- Relazione tra classi di livello 1 e 2
 - la classe di livello 2 può essere un'ampliamento della classe di livello 1
{ liv1: **k, k'** }, { liv2: **k, k', h** }
 - le classi di livello 2 codificano distinzioni ortogonali rispetto a quelle codificate a livello 1
{ liv1: **š, ś** }
{ liv2: **č, čč, ć, š** } { liv2: **s, f, ś, ř, ž, z** } { liv2: **š, šš** }

Criteri adottati nella definizione delle classi di equivalenza

- **Criterio guida:**

- ☞ **ottimizzazione** del rapporto tra la quantità di materiali lessicali recuperati e il rumore prodotto dalla ricerca

- **Conseguenza:**

- ☞ variazioni allofoniche di una certa area corrispondenti ad opposizioni fonematiche di un'altra area sono state escluse dalle classi di equivalenza in quanto avrebbero potuto causare una troppo elevata quantità di rumore nei risultati

Ad esempio:

- lenizione di occlusiva sorda intervocalica: /četrjǫlo/ vs /čedriǫlo/ - /brǫto/ vs /brǫdo/
 - variazione vocalica: /mil'l'áččo/ vs /mil'l'ěččo/ - /álba/ vs /ěrba/

- **Una possibile soluzione:**

- ☞ specializzazione geografica delle classi di equivalenza, o sottoinsiemi di esse, rispetto ad una data sub-area

Interrogazione di livello 0:

forma richiesta:

/skjaččáta/

forme recuperate:

/skjaččáta/

The screenshot shows a window titled 'Quadro' with a search bar containing '1)* {FormD} skjaččáta'. A green circle highlights the search bar, and a yellow circle highlights the number '30' in the top right corner. Below the search bar is a list of results in a window titled 'skiaCCata'. The results are numbered 1) through 9) and include details such as 'Punto', 'TpInc', 'Dom', 'Inf. A', 'Forma', 'CGram', 'SO', 'Descr', 'CUso', and 'CVar'. The search term 'skjaččáta' is highlighted in red in the 'Forma' field of each result.

```
1)* {FormD} skjaččáta 30
```

skiaCCata

- 1) 0140290/1550036 {Punto}014 {TpInc} O {Dom}290 {Inf. A} 1,8
{Forma} <skjaččáta> {CGram} SO
- 2) 0200298/1231642 {Punto}020 {TpInc} O {Dom}298 {Inf. A}
1,2,3,4 {Forma} <skjaččáta> {CGram} SO {Descr} **Torta**
pasquale fatta con la pasta del pane arricchita con olio,
lardo, zucchero, uova, fichi e uvetta. {CUso} AN
- 3) 0280290/1120029 {Punto}028 {TpInc} O {Dom}290 {Inf. A} 1,5
{Forma} <skjaččáta> {CGram} SO
- 4) 0500290/1232057 {Punto}050 {TpInc} O {Dom}290 {Inf. A} 5,7
{Forma} <skjaččáta> {CGram} SO
- 5) 054B290/1130031 {Punto}054 {TpInc} B {Dom}290 {Inf. A}
1,2,3,4,5,6,A,B,C,D {Forma} <skjaččáta> {CGram} SO {Descr}
Grande o piccola. {CRegs} RM(1-4,6,A-C:+)
- 6) 0700290/1150042 {Punto}070 {TpInc} O {Dom}290 {Inf. A}
1,2,3,4,5,6,7,A,B {Forma} <skjaččáta> {CGram} SO {CUso} RE
- 7) 0700298/1150066 {Punto}070 {TpInc} O {Dom}298 {Inf. A}
1,2,3,4,5,6,7,A {Forma} <skjaččáta> {CGram} SO {Descr} **Dolce**
di Pasqua a forma di panettone, con uvetta, semi di anice e
altro. {CUso} DI {CVar} TR
- 8) 0700298/1150067 {Punto}070 {TpInc} O {Dom}298 {Forma}
<skjaččáta> {CGram} SO {Descr} **Torta all' olio, salata.**
{CUso} RE {CVar} NT
- 9) 0710290/1152434 {Punto}071 {TpInc} O {Dom}290 {Inf. A}
1,2,3,4,5,6,7,8 {Forma} <skjaččáta> {CGram} SO {Descr} **Fatta**
con pasta di pane, lavorata bucardola con le dita, unta e
salata in superficie.

per un totale di 30 attestazioni

Interrogazione di livello 1:

forma richiesta:

/skjaččáta/

forme recuperate:

/skjaččáta/

/skjaččáta/

Quadro

1) {FormD}skjaččáta	1
2) {FormD}skjaččáta	30
3) {FormD}skjaččáta	23
4)*{FormD}skjaččáta	1

Famiglia n. 1

- 27) 1780290/1181585 {Punto}178 {TpInc} 0 {Dom}290 {Inf .A} 1,2,3,4,5,6,A,B,C,E {Forma} <skjaččáta> {CGram} SO
- 28) 1920290/1311235 {Punto}192 {TpInc} 0 {Dom}290 {Inf .A} 1,2,3,4,5,6,7 {Forma} <skjaččáta> {CGram} SO
- 29) 1940290/1351118 {Punto}194 {TpInc} 0 {Dom}290 {Inf .A} 5,7 {Forma} <skjaččáta> {CGram} SO
- 30) 1940298/1351154 {Punto}194 {TpInc} 0 {Dom}298 {Inf .A} 1 {Forma} <skjaččáta> {CGram} SO {Descr} Torta pasquale di forma rotonda.
- 31) 2060290/1360072 {Punto}206 {TpInc} 0 {Dom}290 {Inf .A} 3 {Forma} <skjaččáta> {CGram} SO
- 32) 0230290/1210829 {Punto}023 {TpInc} 0 {Dom}290 {Inf .A} 1,2,3,4,5,6,A,B {Forma} <skjaččáta> {CGram} SO
- 33) 0300290/1120685 {Punto}030 {TpInc} 0 {Dom}290 {Inf .A} 9,8 {Forma} <skjaččáta>, <skjaččatina> {CGram} SO
- 34) 0430290/1141193 {Punto}043 {TpInc} 0 {Dom}290 {Inf .A} 3,4,5,6 {Forma} <skjaččáta> {CGram} SO
- 35) 0430298/1141207 {Punto}043 {TpInc} 0 {Dom}298 {Inf .A} 4 {Forma} <skjaččáta> {CGram} SO {Descr} Dolce di Pasqua, fatto con lo zucchero e le uova.
- 36) 0430298/1141208 {Punto}043 {TpInc} 0 {Dom}298 {Inf .A} 5,6 {Forma} <skjaččáta> {CGram} SO
- 37) 0530282/1121552 {Punto}053 {TpInc} 0 {Dom}282 {Inf .A} 3,4 {Forma} <skjaččáta> {CGram} SO
- 38) 0530290/1121569 {Punto}053 {TpInc} 0 {Dom}290 {Inf .A} 1,2,3,4,5,6,7,8 {Forma} <skjaččáta>, <stjaččáta> {CGram} SO {CRegs} RM(+) {Comm} Il termine è di registro non familiare.

per un totale di 55 attestazioni

Interrogazione di livello 2:

forma richiesta:

/skiaččáta/

forme recuperate:

/skiaččáta/

/skiaččáta/

/stiaččáta/

/stiaččáta/

/sčacčáta/

/sčacčáta/

/st'aččáta/

/st'aččáta/

/skiačáta/

...

The screenshot shows a window titled "Ricerca Normalizzata" with a list of 14 search results. Each result is numbered and contains a unique identifier, a set of tags in curly braces, and a form name in angle brackets. Some results include additional information like domain, inflection, and descriptions in red text.

- 1) 0050290/1412224 {Punto}005 {TpInc} O {Dom}290 {Inf.P} 1,6 {Forma} <skjačáta> {CGram} SO {CVar} NT(6:La Spezia;1:tosc)
- 2) 0110290/1541043 {Punto}011 {TpInc} O {Dom}290 {Inf.A} 1,6 {Forma} <stjaččáta> {CGram} SO
- 3) 0110290/1541044 {Punto}011 {TpInc} O {Dom}290 {Inf.A} 3,4,5,9,11 {Forma} <sčacčáta> {CGram} SO
- 4) 0110298/1541070 {Punto}011 {TpInc} O {Dom}298 {Inf.A} 1,5,7 {Forma} <stjaččáta> {CGram} SO {Descr} Torta pasquale a base di uova, zucchero, farina e lievito.
- 5) 0110298/1541071 {Punto}011 {TpInc} O {Dom}298 {Inf.A} 3,5,7 {Forma} <sčacčáta> {CGram} SO {Descr} Torta pasquale a base di uova, zucchero, farina e lievito.
- 6) 0120290/1541559 {Punto}012 {TpInc} O {Dom}290 {Inf.A} 1,2,3,4,5,6 {Forma} <stjaččáta> {CGram} SO
- 7) 0120290/1541560 {Punto}012 {TpInc} O {Dom}290 {Inf.A} 7,8,9,10,11,12 {Forma} <sčacčáta> {CGram} SO
- 8) 0130290/1542002 {Punto}013 {TpInc} O {Dom}290 {Inf.A} 1,2,3,4,5,6,7,8 {Forma} <sčacčáta> {CGram} SO {Descr} Pasta di pane con olio e salvia.
- 9) 0140290/1550035 {Punto}014 {TpInc} O {Dom}290 {Inf.A} 5,6,7 {Inf.P} 8 {Forma} <stjaččáta> {CGram} SO
- 10) 0140290/1550036 {Punto}014 {TpInc} O {Dom}290 {Inf.A} 1,8 {Forma} <skjaččáta> {CGram} SO
- 11) 0150290/1550402 {Punto}015 {TpInc} O {Dom}290 {Inf.A} 2,4,5,6 {Inf.P} 1,3 {Forma} <sčacčáta> {CGram} SO {CUs} RE(1,3)
- 12) 0180290/1551658 {Punto}018 {TpInc} O {Dom}290 {Inf.A} 1 {Forma} <stjaččáta> {CGram} SO
- 13) 0180290/1551659 {Punto}018 {TpInc} O {Dom}290 {Inf.A} 2,3,4,5,6 {Forma} <sčacčáta> {CGram} SO
- 14) 0190290/1202487 {Punto}019 {TpInc} O {Dom}290 {Inf.A} 1,2,3 {Forma} <stjaččáta> {CGram} SO {Descr} Può essere salata o <ššókka>.

per un totale di 309 attestazioni

Conclusioni

La procedura illustrata rappresenta una possibile risposta ad un problema cruciale nell'accesso a dati in trascrizione fonetica non tipizzati:

- la ricerca può astrarre da tratti specifici della realizzazione fonetica del dato attestato
- l'interrogazione avviene secondo livelli differenziati di astrazione incrementale da selezionarsi da parte dell'utente

La costruzione e l'impiego di classi di equivalenza di fonie ai fini dell'interrogazione forniscono:

- una valutazione indiretta della copertura e validità dell'analisi dei dialetti toscani acquisita con l'esperienza dell'ALT
- un punto di partenza nella procedura di tipizzazione manuale dei materiali dialettali raccolti