

# Le tecnologie del linguaggio a sostegno dell'attività didattica

**Giulia Venturi**

Istituto di Linguistica Computazionale «Antonio  
Zampolli» (ILC-CNR)

ItaliaNLP Lab

[www.italianlp.it](http://www.italianlp.it)

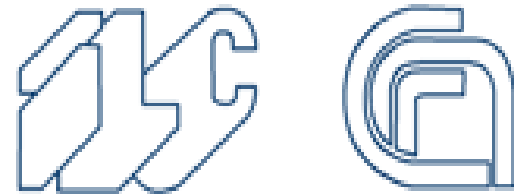
*AIUCD Scuola estiva «Strumenti digitali per umanisti»*

Pisa, 12 giugno 2017

# Italian Natural Language Processing Laboratory @ILC

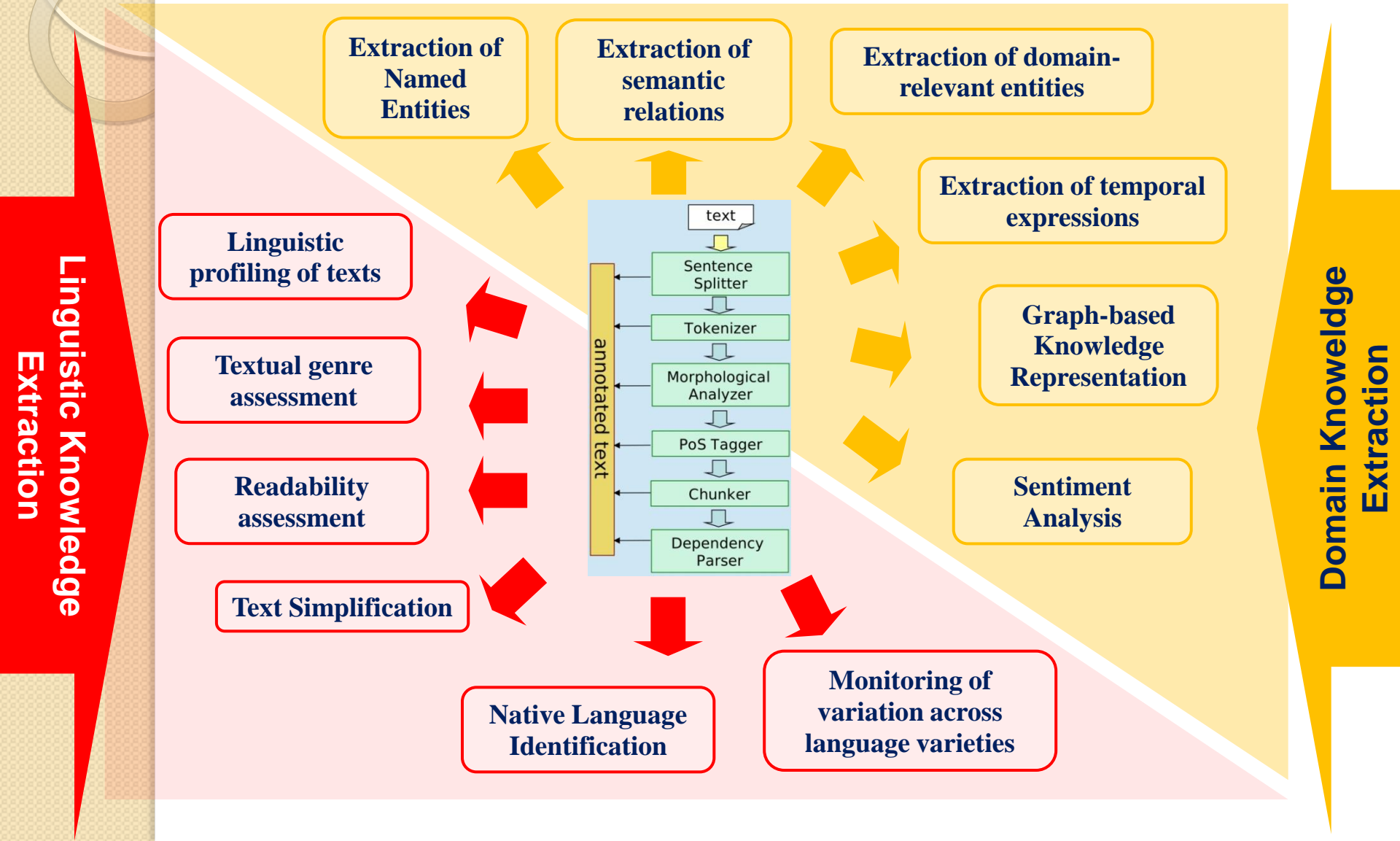
## Persone

Simonetta Montemagni  
Felice Dell'Orletta  
Giulia Venturi  
Andrea Cimino  
Dominique Brunato




[www.italianlp.it](http://www.italianlp.it)

# Le attività di ricerca dell'ItaliaNLP Lab



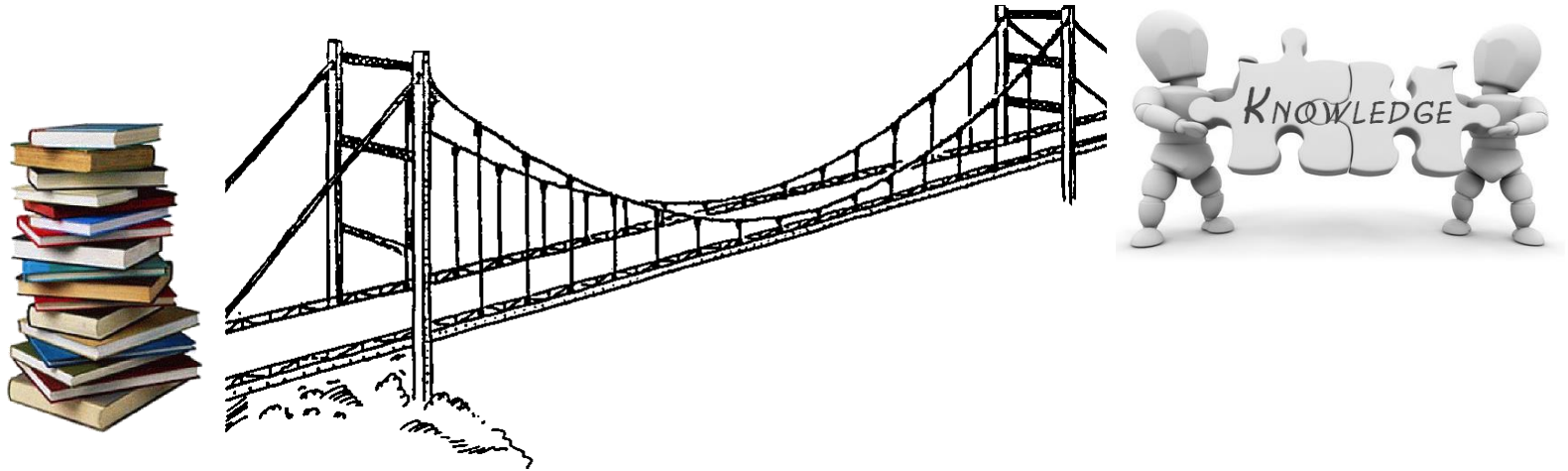
**Linguistic Knowledge Extraction**

**Domain Knowledge Extraction**



Come possono le tecnologie del linguaggio fornire un sostegno per l'attività didattica nella scuola?

# Le tecnologie del linguaggio: cosa sono?



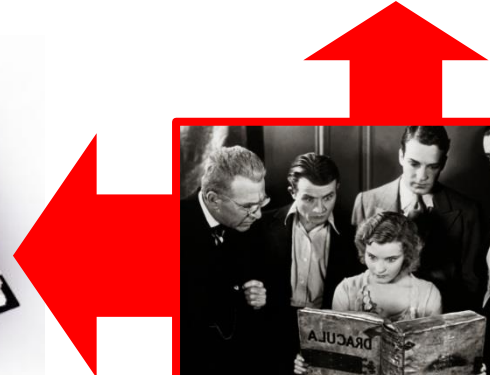
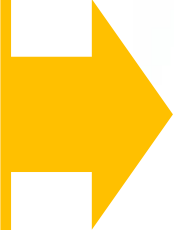
- Sistemi in grado di accedere al contenuto informativo dei testi attraverso l'elaborazione automatica del linguaggio (*Natural Language Processing*)
- Un 'ponte' tra il testo e il contenuto
  - Conoscenza linguistica
    - morfo-sintattico, sintattico, semantico-lessicale
  - Conoscenza di dominio

# Il testo visto dall'uomo e dal computer

```
4C 65 20 61 76 76 65 6E 74 75 72 65 20 64 69 20
50 69 6E 6F 63 63 68 69 6F 6F 0A 18 81 70 89 74
6F 6C 6F 20 49 0D 0A 43 6F 6D 65 20 61 6E 64 F2
20 63 68 65 20 4D 61 65 73 74 72 6F 20 43 69 6C
69 65 67 69 61 2C 20 66 61 6C 65 67 6E 61 6D 65
2C 20 74 72 6F 76 F2 20 75 6E 20 70 65 7A 7A 6F
20 64 69 20 6C 65 67 6E 6F 2C 20 63 68 65 20 70
69 61 6E 67 65 76 61 20 65 20 72 69 64 65 76 61
20 63 6F 6D 65 20 75 6E 20 62 61 6D 62 69 6E 6F
2E 0D 0A 43 27 65 72 61 20 75 6E 61 20 76 6F 6C
74 61 2E 2E 2E 0D 0A 2D 20 55 6E 20 72 65 21 20
2D 20 64 69 72 61 6E 6E 6F 20 73 75 62 69 74 6F
20 61 6E 6E 6E 6E 6E 6E 6E 6E 6E 6E 6E 6E 6E 6E
6C 65 74 74 6F 72 69 2E 0D 0A 2D 20 4E 6F 2C 20
72 61 67 61 7A 7A 69 2C 20 61 76 65 74 65 20 73
62 61 67 6C 69 61 74 6F 2E 20 43 27 65 72 61 20
75 6E 61 20 76 6F 6C 74 61 20 75 6E 20 70 65 7A
7A 6F 20 64 69 20 6C 65 67 6E 6F 2E 0D 0A 4E 6F
6E 20 65 72 61 20 75 6E 20 6C 65 67 6E 6F 20 64
69 20 6C 75 73 73 6F 2C 20 6D 61 20 75 6E 20 73
65 6D 70 6C 69 63 65 20 70 65 7A 7A 6F 20 64 61
20 63 61 74 61 73 74 61 2C 20 64 69 20 71 75 65
6C 6C 69 20 63 68 65 20 64 27 69 6E 76 65 72 6E
6F 20 73 69 20 6D 65 74 74 6F 6E 6F 20 6E 65 6C
6C 65 20 73 74 75 66 65 20 65 20 6E 65 69 20 63
61 6D 69 6E 65 74 74 69 20 70 65 72 20 61 63 63
65 6E 64 65 72 65 20 69 6C 20 66 75 6F 63 6F 20
65 20 70 65 72 20 72 69 73 63 61 6C 64 61 72 65
20 6C 65 20 73 74 61 6E 7A 65 2E 0D 0A 4E 6F 6E
20 73 6F 20 63 6F 6D 65 20 61 6E 64 61 73 73 65
```

Le avventure di Pinocchio. Capitolo I... Come andò che Maestro Ciliegia, falegname, trovò un pezzo di legno, che piangeva e rideva come un bambino... C'era una volta.....- Un re! Diranno subito i miei piccoli lettori...- No, ragazzi, avete sbagliato. C'era una volta un pezzo di legno... Non era un legno di lusso, ma un semplice pezzo da catasta, di quelli che d'inverno si mettono nelle stufe e nei caminetti per accendere il fuoco e per riscaldare le stanze... Non so come andasse

Non tutti guardano le cose allo stesso modo



# Le tecnologie del linguaggio: l'annotazione linguistica

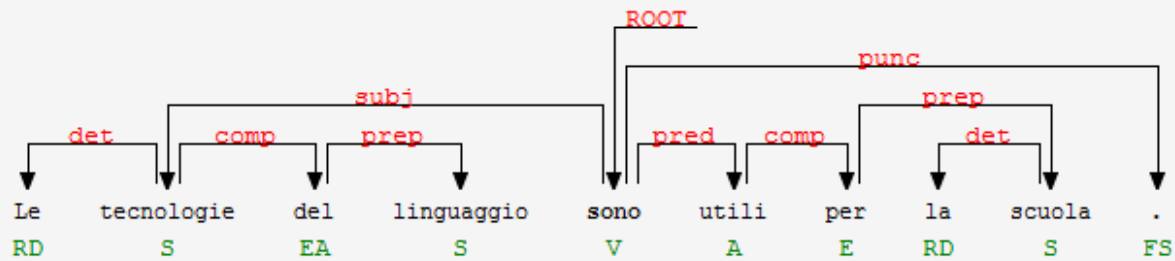
- L'annotazione linguistica consiste nella codifica di informazione linguistica associata al testo
  - Rende interpretabile dal computer la struttura linguistica implicita nel testo
- Catena di analisi linguistica
  - Sviluppata da ILC e Università di Pisa
  - <http://www.italianlp.it/demo/linguistic-annotation-tool/>

# L'annotazione linguistica

## Linguistic Annotation Pipeline

[www.italianlp.it](http://www.italianlp.it)

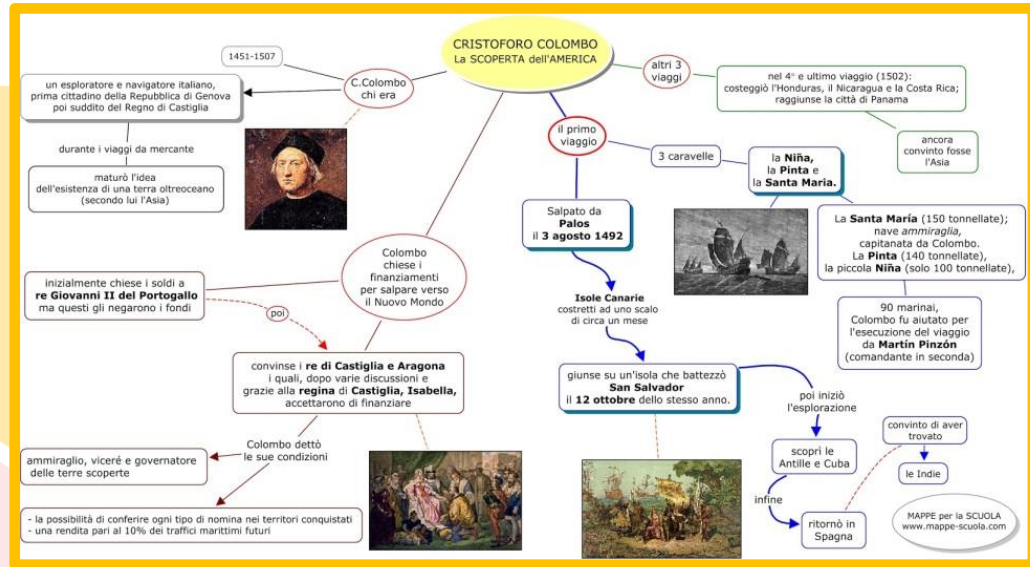
Text ✓ Sentence Splitting ✓ Part of Speech Tagging ✓ Syntactic Parsing ✓ Syntactic Trees ✓





# Le tecnologie del linguaggio per la didattica

Guida alla costruzione di mappe concettuali dai testi



## Analisi delle caratteristiche linguistiche

- delle produzioni scritte degli apprendenti e
- dei materiali didattici

L'accesso alla struttura linguistica

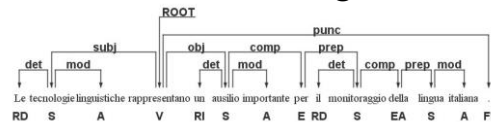
L'accesso al contenuto

# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

- La metodologia:



Annotazione linguistica



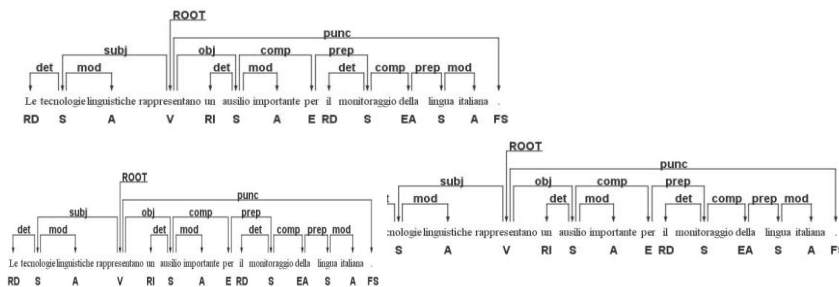
**MONITOR-It**

Strumento per l'estrazione automatica delle caratteristiche linguistiche di un testo

Analisi della struttura linguistica per il monitoraggio dell'evoluzione nel tempo delle abilità linguistiche

# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

Corpus linguisticamente analizzato in modo automatico

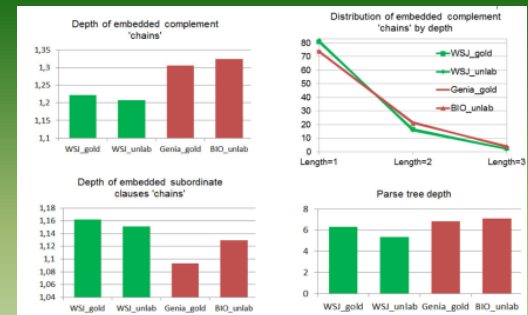


Vocabolario di Base

**MONITOR-It**

Strumento per l'estrazione automatica delle caratteristiche linguistiche di un testo

- Il profilo linguistico del testo è ricostruito sulla base delle caratteristiche linguistiche rilevate rispetto a diversi livelli di annotazione linguistica
  - Lemmatizzazione
  - Annotazione morfosintattica
  - Annotazione sintattica a dipendenze

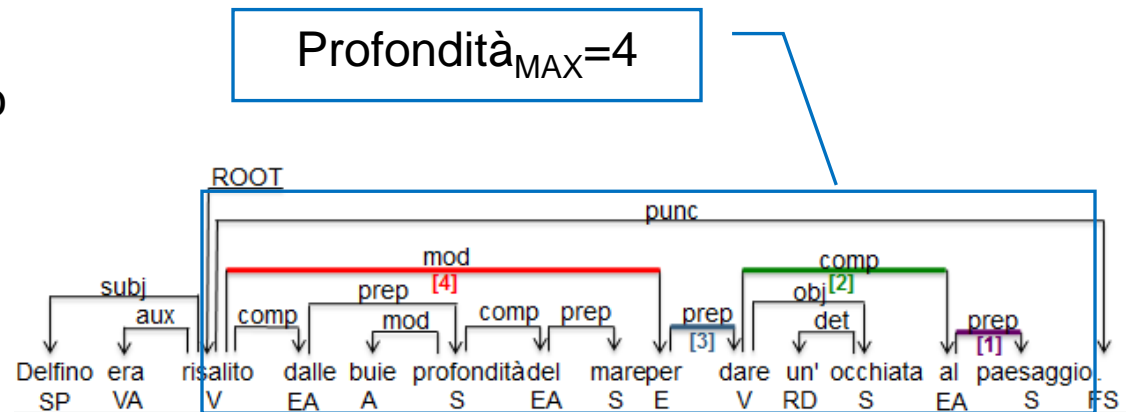


# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

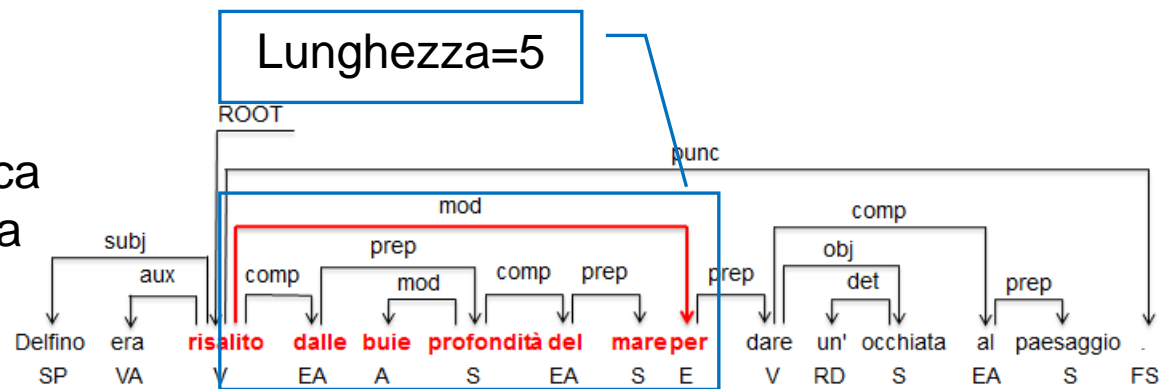
Generali	Lunghezza media delle frasi e delle parole
Lessicali	Lemmi riconducibili al Vocabolario di Base del GRADIT (De Mauro, 2000)
	Indice di ricchezza lessicale (Type/token ratio)
Morfosintattiche	Distribuzione delle categorie morfosintattiche
	Densità lessicale (parole piene/funzionali)
Sintattiche	Distribuzione delle relazioni di dipendenza sintattica
	Profondità dell'albero sintattico (numero massimo di relazioni di dipendenza consecutive)
	Lunghezza delle relazioni di dipendenza sintattica, calcolata come la distanza tra la testa e il dipendente (in tokens)
	Lunghezza media di catene di dipendenza sintattica a testa nominale
	Valenza verbale, calcolata come il numero delle dipendenze effettive di una testa verbale
	Rapporto tra frasi principali e subordinate
	Lunghezza media di sequenze di subordinate consecutive
	Lunghezza media delle relazioni di dipendenza, calcolata in numero di tokens consecutivi (es. distanza soggetto/verbo)

# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

- Profondità dell'albero sintattico (numero massimo di relazioni di dipendenza consecutive)



- Lunghezza delle relazioni di dipendenza sintattica (distanza tra la testa e il dipendente, in tokens)



# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

- Un *case study* (Barbagli et al., 2014):
  - Studio dell'evoluzione delle abilità di scrittura dalla prima alla seconda classe della scuola secondaria di primo grado
- Obiettivo:
  - **Pedagogia sperimentale**: costruzione di un modello di analisi empirica in grado di consentire l'osservazione dei processi e dei prodotti della didattica della produzione scritta
  - **Linguistica computazionale**: definizione di un modello computazionale in grado di intercettare le caratteristiche linguistiche maggiormente coinvolte nel processo di evoluzione

# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

- Il corpus

7 scuole di due aree territoriali collocate agli estremi della suddivisione socio-culturale ed economica di Roma: **3** scuole del centro e **4** scuole di periferia

CENTRO	n. alunni	PERIFERIA	n. alunni
Scuola A	26	Scuola D	13
Scuola B	27	Scuola E	24
Scuola C	24	Scuola F	20
		Scuola G	22
<b>Totale</b>	<b>77</b>		<b>79</b>
<b>Totale elaborati: 1.352</b>			

## Questionario studenti: esempi

		Centro	Periferia
Titolo di studio della madre	Entro l'obbligo	3,2	50,8
	Diploma	19,4	26,2
	Laurea	77,4	23,1
	<b>Totale</b>	<b>100,0</b>	<b>100,0</b>
Titolo di studio del padre	Entro l'obbligo	6,7	53,3
	Diploma	16,7	33,3
	Laurea	76,7	13,3
	<b>Totale</b>	<b>100,0</b>	<b>100,0</b>

		Centro	Periferia
Lavoro della madre	Operai e artigiani	18,3	68,2
	Medie professionalità	26,8	25,8
	Alte professionalità	54,9	6,1
	<b>Totale validi</b>	<b>100</b>	<b>100</b>
Lavoro del padre	Operai e artigiani	16,2	75,7
	Medie professionalità	36,5	21,4
	Alte professionalità	47,3	2,9
	<b>Totale validi</b>	<b>100</b>	<b>100</b>

# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

- Due tipi di annotazione:
  - Annotazione linguistica (automatica)
  - Annotazione e correzione degli errori commessi dagli apprendenti (manuale)
- Definizione di uno schema di annotazione degli errori oggi non esistente per la lingua italiana
  - Organizzata su tre livelli:
    - aree di competenza linguistica: grammatica, ortografia, lessico
    - categoria morfo-sintattica coinvolta, es. verbo, articolo, pronome, preposizione
    - tipo di errore, es. omissione, eccesso, mancanza di concordanza
  - Esempio di annotazione di uso errato del pronome relativo:
    - *La scienza non so perché mi fa pensare a un fenomeno costruito su un'altura <M t="19" c="per cui">che</M>ci vuole molto ingegno.*



# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

Categoria	Sotto-categoria	I anno			II anno			Totale %
		Freq.%	Media	Dev.	Freq.%	Media	Dev.	
<b>Grammatica</b>								
Verbi	Uso dei tempi	7,78 (150)	0,99	2,29	15,67 (239)	1,47	4,05	11,26 (389)
	Uso dei modi	4,25 (82)	0,54	1,39	4,92 (75)	0,49	0,99	4,55 (157)
	Concordanza con il soggetto	2,85 (55)	0,37	1,38	4 (61)	0,41	1,27	3,36 (116)
Preposizioni	Uso errato	6,48 (125)	0,83	2,58	6,75 (103)	0,66	1,21	6,6 (228)
	Omissione o eccesso	1,03 (20)	0,13	0,40	0,72 (11)	0,07	0,25	0,90 (31)
Pronomi	Uso errato	5,09 (98)	0,65	1,13	3,54 (54)	0,36	0,97	4,4 (152)
	Omissione	0,41 (8)	0,05	0,36	0,59 (9)	0,06	0,39	0,49 (17)
	Eccesso	2,70 (52)	0,35	0,61	1,57 (24)	0,16	0,46	2,2 (76)
	Uso errato del pronome relativo	2,13 (41)	0,27	0,70	1,70 (26)	0,17	0,44	1,94 (67)
Articoli	Uso errato	5,81 (112)	0,75	3,72	3,54 (54)	0,35	1,09	4,81 (166)
Congiunzioni e/o connettivi	Uso errato	0,57 (11)	0,07	0,33	0,52 (8)	0,05	0,23	0,55 (19)
Altro		7,31 (141)	0,94	3,66	5,18 (79)	0,49	1,79	6,37 (220)
<b>Ortografia</b>								
Doppie	Uso per difetto	6,74 (130)	0,83	2,49	5,05 (77)	0,48	1,56	5,99 (207)
	Eccesso	3,27 (63)	0,42	0,89	3,67 (56)	0,37	1,13	3,45 (119)
Uso dell'h	Per difetto	3,21 (62)	0,39	1,03	1,64 (25)	0,17	0,62	2,52 (87)
	Per eccesso	1,66 (32)	0,21		1,11 (17)	0,10		1,42 (49)
Monosillabi	Uso errato dei monosillabi accentati	4,87 (94)	0,63	1,07	4,07 (62)	0,40	0,83	4,52 (156)
	Uso di <i>po</i> o <i>pò</i> anziché <i>po'</i>	1,66 (32)	0,21	0,72	1,64 (25)	0,17	0,52	1,65 (57)
Apostrofo	Uso errato	4,82 (93)	0,61	1,01	4,52 (69)	0,46	0,89	4,69 (162)
Altro		21,77 (420)	2,76	4,58	23,02 (351)	2,27	4,60	22,32 (771)
<b>Lessico</b>								
Vocabolario	Uso errato	5,60 (108)	0,70	1,64	6,56 (100)	0,66	1,09	6,02 (208)
Numero totale di errori		1929			1525			

# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

- Due tipologie di risultati:
  - ricostruzione del **profilo linguistico** delle produzioni rispetto
  - ricostruzione del **processo di apprendimento** delle abilità di scrittura
- Due tipi di prospettiva di analisi dei risultati:
  - rispetto al mutamento diacronico
  - rispetto alle variabili geografiche e di sfondo del questionario studenti

# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

- Ricostruzione del **profilo linguistico** delle produzioni

## Mutamento diacronico

Caratteristiche linguistiche	I anno (%)	II anno (%)
Preposizioni	10,68	11,37
Preposizioni articolate	2,82	3,27
Pronomi	9,31	8,38
Pronomi clitici	4,79	4,32
Pronomi personali	1,7	1,27
Sostantivi	19,92	20,66

## Mutamento diacronico e variabile geografica

Area urbana		Lunghezza media relazioni di dipendenza	Altezza massima degli alberi sintattici
Centro	I anno	2,37	6,05
	II anno	2,42	6,20
Periferia	I anno	2,55	6,69
	II anno	2,47	6,84

# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

- Identificazione della lingua madre a partire da produzioni di apprendenti L2
  - **Native Language Identification**: *Classificazione automatica di produzioni linguistiche di apprendenti L2 sulla base della L1* (Koppel et al., 2005)
- Un *case study* (Cimino et al., 2013):
  - 12.100 prove TOEFL
  - 11 lingue L1: arabo, cinese, francese, tedesco, hindi, italiano, giapponese, coreano, spagnolo, telugu, turco
  - 3 livelli: high, medium, low
  - 8 tracce (es. «Do you agree or disagree with the following statement? The best way to travel is in a group led by a tour guide. Use reasons and examples to support your answer.»)
- Risultati: 78% di accuratezza
- Ricaduta applicativa:
  - supporto all'insegnante nella personalizzazione della sua azione educativa

# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

- Ricostruzione del **processo di apprendimento** delle abilità di scrittura
- Ipotesi di partenza:
  - data una coppia di temi di uno studente, ordinata cronologicamente, il secondo documento dovrebbe presentare una forma linguistica più evoluta rispetto al precedente
- Il processo di apprendimento è trattato come un compito di classificazione probabilistica:
  - per ogni studente e sulla base delle caratteristiche linguistiche estratte, il classificatore dovrà predire l'ordine cronologico di scrittura delle coppie di temi attribuendo classe
    - 1: se  $t(d2) > t(d1)$  [d2 è stato scritto DOPO d1]
    - 0: se  $t(d2) < t(d1)$  [d2 è stato scritto PRIMA di d1]

# Analisi delle caratteristiche linguistiche delle produzioni scritte di apprendenti

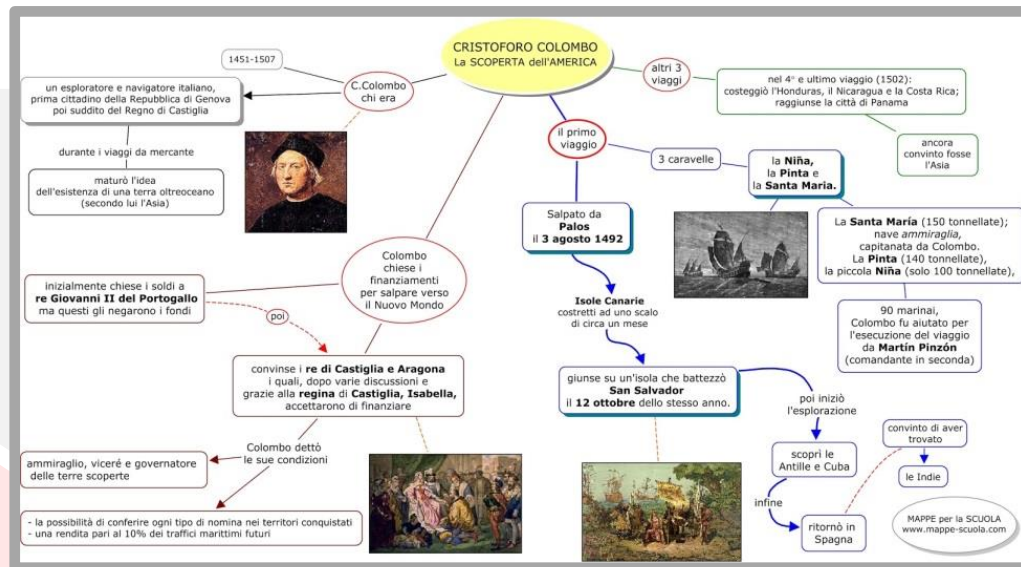
- Ricostruzione del **processo di apprendimento** delle abilità di scrittura
- Qualche risultato
  - *Alessio Miaschi, 4 maggio 2016, tesi magistrale in Informatica Umanistica*

Intervallo temporale	F-score
I anno - I anno (senza prova comune)	0.68
I anno - I anno (con la prova comune)	0.73
Prova comune I anno – I anno	0.96

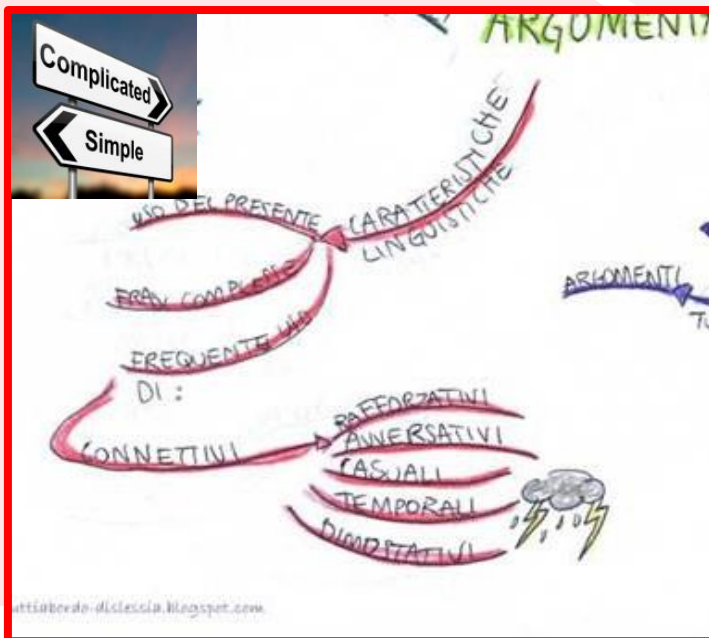
- Alcune delle caratteristiche linguistiche più determinanti:
  - caratteristiche di base del testo (lunghezza frase)
  - uso di forme verbali
  - uso dei pronomi
  - posizione del soggetto pre e post-verbale

# Le tecnologie del linguaggio per la didattica

Guida alla costruzione di mappe concettuali dai testi



L'accesso alla struttura linguistica



## Analisi delle caratteristiche linguistiche

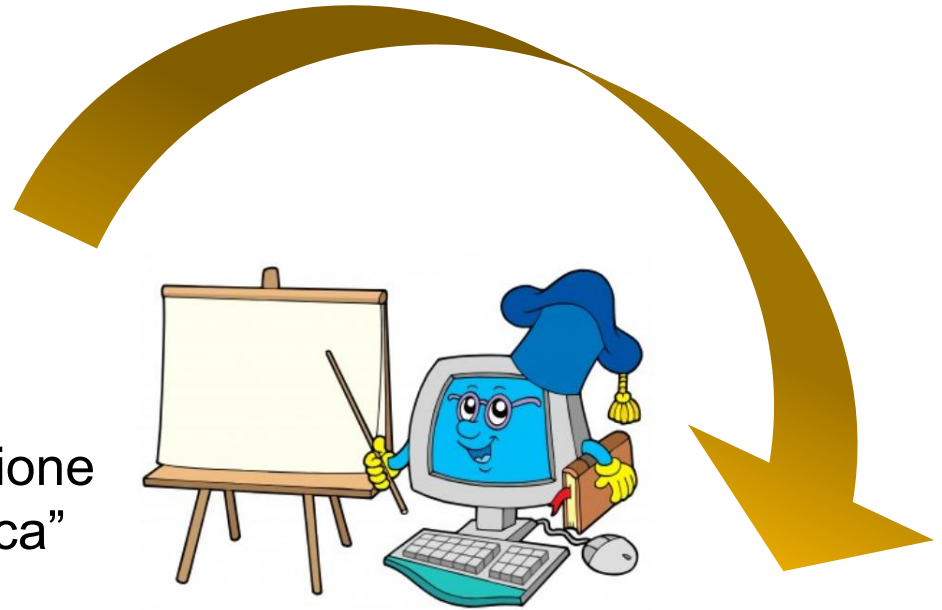
- delle produzioni scritte degli apprendenti e
- dei materiali didattici

L'accesso al contenuto

# Ausilio nella **scrittura e personalizzazione dei materiali didattici**



Nelle Linee Guida dell’Azione “Editoria Digitale Scolastica” emanata dal MIUR si prefigurano “prodotti multimediali le cui singole componenti possano essere utilizzate dai docenti per lo **sviluppo di materiali didattici personalizzati**”





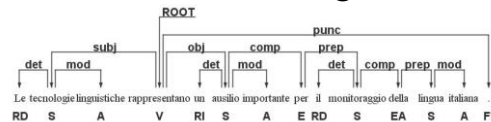
# Analisi delle caratteristiche linguistiche dei materiali didattici

- La metodologia:

## MATERIALI DIDATTICI



## Annotazione linguistica



## READ-IT

Strumento automatico per l'analisi della leggibilità di un testo in italiano



Valutazione della leggibilità e individuazione dei luoghi di complessità del testo in vista della loro semplificazione

# READ-IT

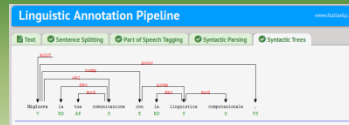
strumento automatico per l'analisi della leggibilità di un testo in italiano

Monitoraggio delle caratteristiche linguistiche di collezioni di testi

Studio dei fattori che rendono un testo complesso

Modelli della comprensione linguistica

## Analisi linguistica



Valutazione dell'**efficacia comunicativa** di testi nella comunicazione

- **Insegnante-studente**
- **Amministratore-Cittadino**
- **Operatore di Call Center-utente finale**
- **Medico-Paziente**
- **Autore editoria scolastica-studenti**

# Un ausilio alla scrittura semplificata del testo

Supporto alla (ri)scrittura personalizzata di materiali mirati alle necessità della classe e/o del singolo



Riscrittura del testo seguendo le indicazioni di READ-IT



# La valutazione della leggibilità: esempi

Versione originale

Testo da analizzare	<u>Suddivisione in frasi</u>	<u>Suddivisione in token</u>	<u>Parti del discorso</u>	<u>Annotazione</u>	<u>Analisi globale della leggibilità</u>	<u>Proiezione della leggibilità sul testo</u>
---------------------	------------------------------	------------------------------	---------------------------	--------------------	--	---

Nota: ogni ritorno a capo costituisce un'interruzione di frase. Prestare attenzione ai testi copiati da documenti pdf.

Il Futurismo in Italia

L'affacciarsi sulla scena culturale europea del Futurismo, prima avanguardia artistica in Italia, è un evento determinante anche nel modo di concepire l'arte e la figura dell'artista. Nella nuova ottica futurista, l'arte non deve più essere un'attività contemplativa confinata nelle accademie e nei musei, ma l'esercizio di una forza vitale operante all'interno della società; e l'artista, in quanto artefice di tale esercizio, deve acquisire un ruolo sociale nuovo e una partecipazione diretta alla dimensione della storia.

Questa aspirazione a far coincidere l'arte con la vita, che è il Futurismo a introdurre, diviene un tratto tipico delle avanguardie artistiche d'inizio Novecento.

L'arte, per i Futuristi, è azione, è forza perturbante la realtà. E della realtà, l'arte futurista, investe tutti gli aspetti: oltre alle arti figurative, anche la moda, la cucina, la fotografia, il cinema, la musica, l'architettura, il teatro, la letteratura (48), la grafica pubblicitaria ed editoriale.

Carattere tipico del Futurismo è infatti l'utilizzo di tutti i possibili canali di comunicazione, per una diffusione più ampia possibile della propria poetica.

Il Futurismo è anche la prima avanguardia artistica che rediga manifesti d'intenti programmatici.

Il primo, pubblicato nel 1909 sul quotidiano parigino «Le Figaro» dal poeta e scrittore italiano Filippo Tommaso Marinetti (1876-1944), attesta la nascita ufficiale di questo movimento d'avanguardia.

Elimina il testo

Analisi completata: i risultati dell'analisi sono ora disponibili nelle apposite schede.

Il testo è nella seguente lingua: ITALIANO ▾ Avvia l'analisi del testo...

# La valutazione della leggibilità: esempi

Versione originale

Testo da analizzare	Suddivisione in frasi	Suddivisione in token	Parti del discorso	Annotazione	Analisi globale della leggibilità	Proiezione della leggibilità sul testo
<b>indice di leggibilità</b>		<b>livello di difficoltà</b>				
READ-IT Base				97,5%		
READ-IT Lessicale				84,8%		
READ-IT Sintattico				77,0%		
READ-IT Globale				99,6%		
<b>indice di leggibilità</b>		<b>livello di semplicità</b>				
GULPEASE				43,3		
<b>[+] [-] Caratteristiche estratte dal testo</b>						
<b>[+] Profilo di base</b>						
<b>[+] Profilo lessicale</b>						
<b>[+] Profilo sintattico</b>						

Alta leggibilità



Bassa leggibilità

# La valutazione della leggibilità: esempi

Testo da analizzare	<u>Suddivisione in frasi</u>	<u>Suddivisione in token</u>	<u>Parti del discorso</u>	<u>Annotazione</u>	<u>Analisi globale della leggibilità</u>	<u>Proiezione della leggibilità sul testo</u>
<p><b>Nota: ogni ritorno a capo costituisce un'interruzione di frase. Prestare attenzione ai testi copiati da documenti pdf.</b></p>						
<p>Accanto all'esperienza cubista e avendo con essa molti punti in comune nasce negli stessi anni, a opera di artisti italiani, il Futurismo.</p> <p>Nel 1909, a Parigi, Filippo Tommaso Marinetti, pubblica sul giornale "Le Figaro" il primo Manifesto del Futurismo. L'anno dopo a Milano si raduna il primo gruppo di artisti futuristi; la loro esperienza artistica non si limita alla pittura e alla scultura, ma spazia anche nell'ambito dell'architettura, della letteratura, del teatro, della poesia, della musica.</p> <p>Ai futuristi interessano i progressi della scienza e le innovazioni tecnologiche, "magnificare la vita odierna, la bellezza della velocità".</p> <p>Proprio perché la realtà è in continuo movimento non esiste più un unico punto d'osservazione, l'oggetto e lo spazio circostante sembrano fondersi, la prospettiva scompare.</p> <p>Pittori futuristi sono: Umberto Boccioni, Carlo Carrà, Giacomo Balla, Gino Severini, Luigi Russolo, Fortunato Depero.</p> <p>A Boccioni non interessa riprodurre la semplice sensazione del movimento relativo, bensì rendere il dinamismo universale, o assoluto, che è dentro alle forme stesse.</p> <p>La sua idea è quella di una forma in continua espansione nello spazio.</p> <p>A Giacomo Balla invece interessa il movimento come visto in una ripresa cinematografica: come in un susseguirsi di fotogrammi sono presenti simultaneamente diverse immagini che evidenziano la dinamicità della scena rappresentata.</p>						
						<input type="button" value="Cancella il testo"/>
<p>Analisi completata: i risultati dell'analisi sono ora disponibili nelle apposite schede.</p>						
<p>Il testo è nella seguente lingua: ITALIANO ▾ <input type="button" value="Avvia l'analisi del testo..."/></p>						

# La valutazione della leggibilità: esempi

Versione semplificata

Testo da analizzare	Suddivisione in frasi	Suddivisione in token	Parti del discorso	Annotazione	Analisi globale della leggibilità	Proiezione della leggibilità sul testo
<b>indice di leggibilità</b>		<b>livello di difficoltà</b>				
READ-IT Base		86,6%				
READ-IT Lessicale		94,7%				
READ-IT Sintattico		4,9%				
READ-IT Globale		50,0%				
<b>indice di leggibilità</b>		<b>livello di semplicità</b>				
GULPEASE		46,3				
<b>[+] [-] Caratteristiche estratte dal testo</b>						
[+] Profilo di base						
[+] Profilo lessicale						
[+] Profilo sintattico						

Alta leggibilità



Bassa leggibilità

# La valutazione dei luoghi di complessità linguistica

## Versione originale

SID	frase	base	less.	sint.	glob.
1.	Il Futurismo in Italia				
2.	L'affacciarsi sulla scena culturale europea del Futurismo, prima avanguardia artistica in Italia, è un evento determinante anche nel modo di concepire l'arte e la figura dell'artista.				
3.	Nella nuova ottica futurista, l'arte non deve più essere un'attività contemplativa confinata nelle accademie e nei musei, ma l'esercizio di una forza vitale operante all'interno della società; e l'artista, in quanto artefice di tale esercizio, deve acquisire un ruolo sociale nuovo e una partecipazione diretta alla dimensione della storia.				
4.	Questa aspirazione a far coincidere l'arte con la vita, che è il Futurismo a introdurre, diviene un tratto tipico delle avanguardie artistiche d'inizio Novecento.				
5.	L'arte, per i Futuristi, è azione, è forza perturbante la realtà.				
6.	E della realtà, l'arte futurista, investe tutti gli aspetti: oltre alle arti figurative, anche la moda, la cucina, la fotografia, il cinema, la musica, l'architettura, il teatro, la letteratura (48), la grafica pubblicitaria ed editoriale.				
7.	Carattere tipico del Futurismo è infatti l'utilizzo di tutti i possibili canali di comunicazione, per una diffusione più ampia possibile della propria poetica.				
8.	Il Futurismo è anche la prima avanguardia artistica che rediga manifesti d'intenti programmatici.				
9.	Il primo, pubblicato nel 1909 sul quotidiano parigino «Le Figaro» dal poeta e scrittore italiano Filippo Tommaso Marinetti (1876-1944), attesta la nascita ufficiale di questo movimento d'avanguardia.				
10.	Vi sono celebrati lo sprezzo del pericolo, il coraggio fisico, la ribellione; vi è proclamata la bellezza della velocità, che ha portato a un radicale mutamento nella percezione della realtà («Un automobile ruggente, che sembra correre sulla mitraglia, è più bello della Vittoria di Samotracia»), l'importanza dell'aggressività («Nessuna opera che non abbia un carattere aggressivo può essere un capolavoro») e la necessità della guerra («sola igiene del mondo») e dell'affrancamento dalla tradizione.				
11.	Il Manifesto tecnico della pittura futurista				

## Versione semplificata

SID	frase	base	less.	sint.	glob.
1.	Accanto all'esperienza cubista e avendo con essa molti punti in comune nasce negli stessi anni, a opera di artisti italiani, il Futurismo.				
2.	Nel 1909, a Parigi, Filippo Tommaso Marinetti, pubblica sul giornale "Le Figaro" il primo Manifesto del Futurismo.				
3.	L'anno dopo a Milano si raduna il primo gruppo di artisti futuristi; la loro esperienza artistica non si limita alla pittura e alla scultura, ma spazia anche nell'ambito dell'architettura, della letteratura, del teatro, della poesia, della musica.				
4.	Ai futuristi interessano i progressi della scienza e le innovazioni tecnologiche, "magnificare la vita odierna, la bellezza della velocità".				
5.	Proprio perché la realtà è in continuo movimento non esiste più un unico punto d'osservazione, l'oggetto e lo spazio circostante sembrano fondersi, la prospettiva scompare.				
6.	Pittori futuristi sono: Umberto Boccioni, Carlo Carrà, Giacomo Balla, Gino Severini, Luigi Russolo, Fortunato Depero.				
7.	A Boccioni non interessa riprodurre la semplice sensazione del movimento relativo, bensì rendere il dinamismo universale, o assoluto, che è dentro alle forme stesse.				
8.	La sua idea è quella di una forma in continua espansione nello spazio.				
9.	A Giacomo Balla invece interessa il movimento come visto in una ripresa cinematografica: come in un susseguirsi di fotogrammi sono presenti simultaneamente diverse immagini che evidenziano la dinamicità della scena rappresentata.				



# Riscrittura del testo per chi?

- Supporto alla (ri)scrittura personalizzata di materiali mirati alle necessità della classe e/o del singolo
- Quanto sono distanti i materiali didattici dalle **diverse** competenze linguistiche degli apprendenti?



Distanza:  
Ortografia  
Lessico  
Morfologia  
Sintassi  
Semantica



# Distanza tra la «norma» e la competenza degli apprendenti

- La differenza tra le caratteristiche linguistiche degli elaborati degli studenti rispetto ai materiali a cui sono esposti è indicativa della distanza tra la competenza linguistica degli apprendenti e la “norma” di riferimento

# Distanza tra la «norma» e la competenza degli apprendenti

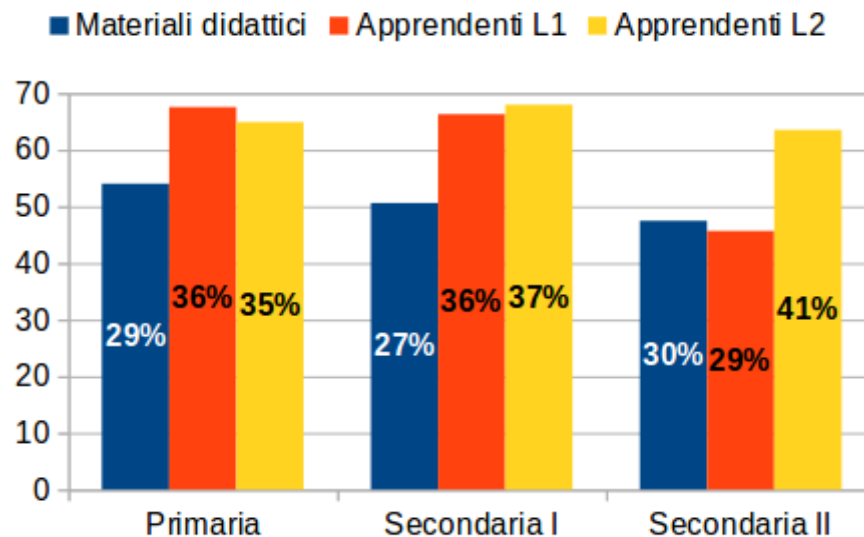
- La differenza tra le caratteristiche linguistiche degli elaborati degli studenti rispetto ai materiali a cui sono esposti è indicativa della distanza tra la competenza linguistica degli apprendenti e la “norma” di riferimento

- Un esempio:
  - la distribuzione delle parole contenute nel Vocabolario di Base (De Mauro, 2000)

Tipologia	Descrizione	Livello scolastico	N. parole
Elaborati di apprendenti italiano L1	Giornalini degli studenti, raccolte di storie scritte da bambini, diari di viaggio dagli studenti, forum di discorsi di classe ...	Scuola primaria	17.939
		Scuola secondaria di primo grado	19.605
		Scuola secondaria di secondo grado	47.237
Materiali didattici distribuiti nella scuola	Libri di testo, sussidiari, modelli di prove scritte per l'esame di stato, pratiche e esercizi didattici, ...	Scuola primaria	47.993
		Scuola secondaria di primo grado	65.260
		Scuola secondaria di secondo grado	48.039
Elaborati di apprendenti italiano L2	Elaborati sul tema “L'Italia è un paese in cui, oltre agli italiani, vivono molti stranieri. Racconta la tua esperienza”	Scuola primaria	2.615
		Scuola secondaria di primo grado	1.286
		Scuola secondaria di secondo grado	2.854

# Distanza tra la «norma» e la competenza degli apprendenti

Distribuzione % delle parole rispetto al Vocabolario di Base



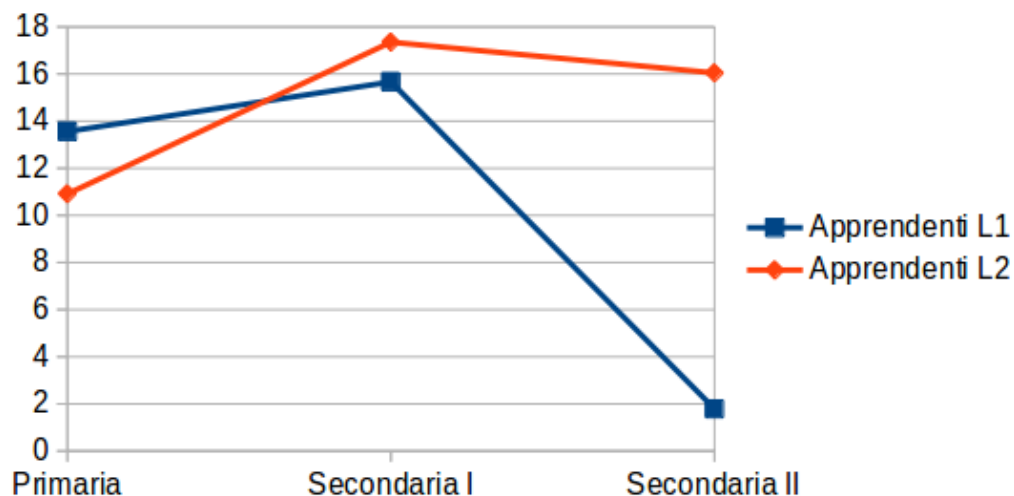
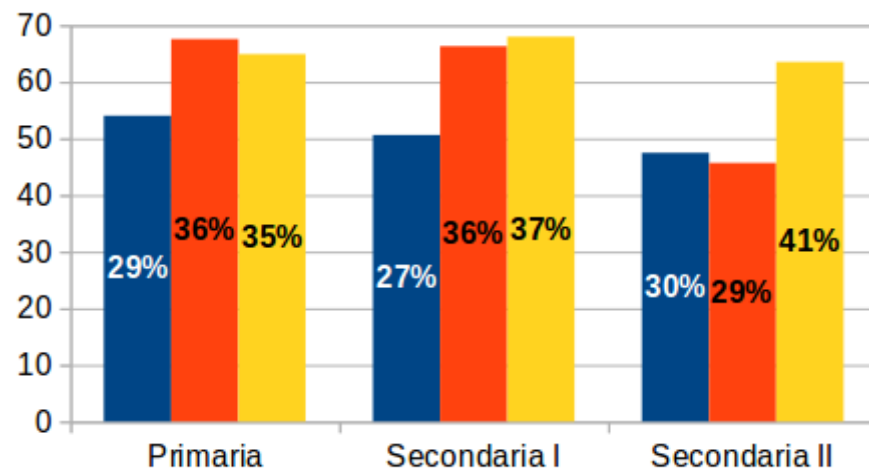
# Distanza tra la «norma» e la competenza degli apprendenti

Distribuzione % delle parole rispetto al Vocabolario di Base

C'è una distanza maggiore tra la competenza lessicale degli apprendenti L2 e i materiali didattici rispetto alla competenza degli L1

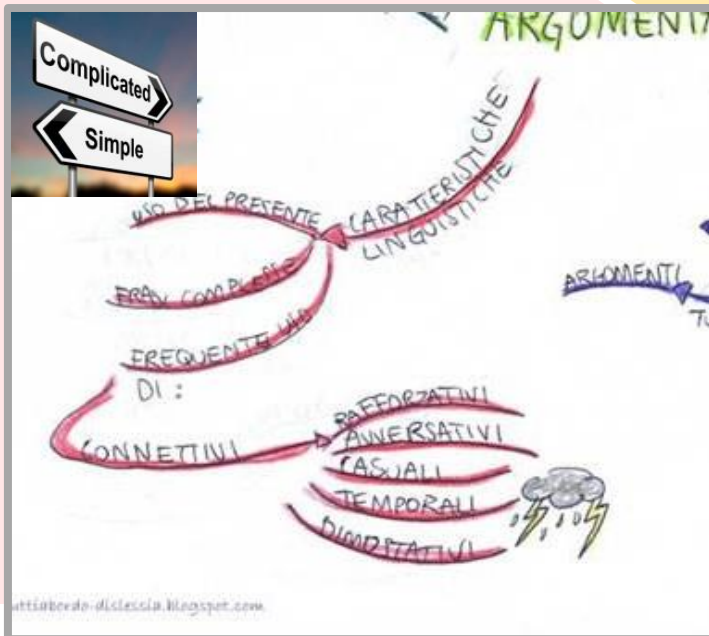
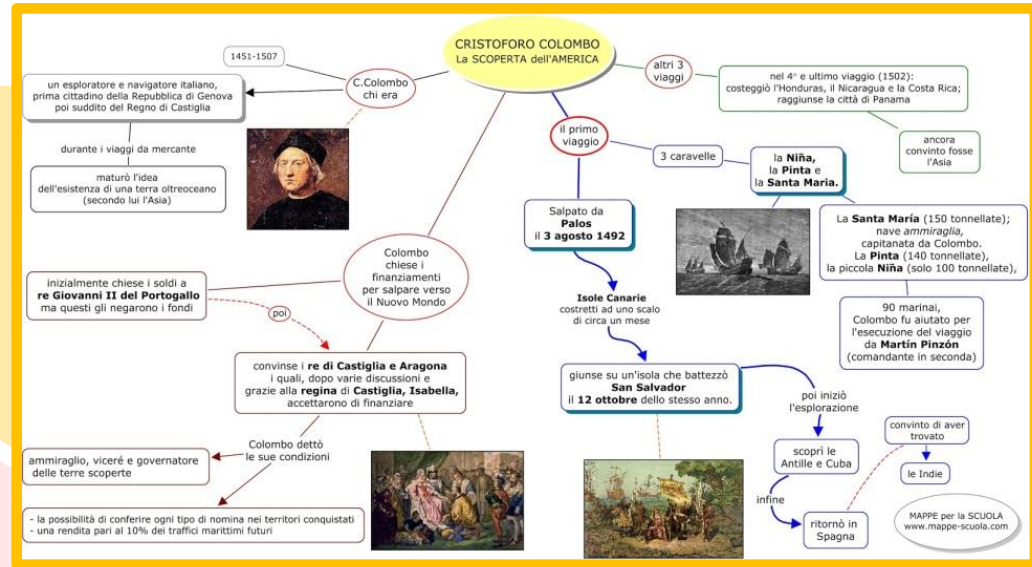
Differenza tra i materiali didattici e le produzioni rispetto alla % di parole del Vocabolario di Base

■ Materiali didattici ■ Apprendenti L1 ■ Apprendenti L2



# Le tecnologie del linguaggio per la scuola

Guida alla costruzione di mappe concettuali dai testi

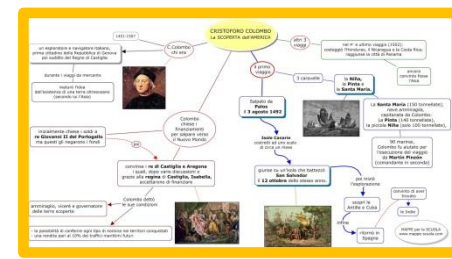


Analisi delle caratteristiche linguistiche dei testi per la valutazione della loro leggibilità e il monitoraggio della distanza linguistica

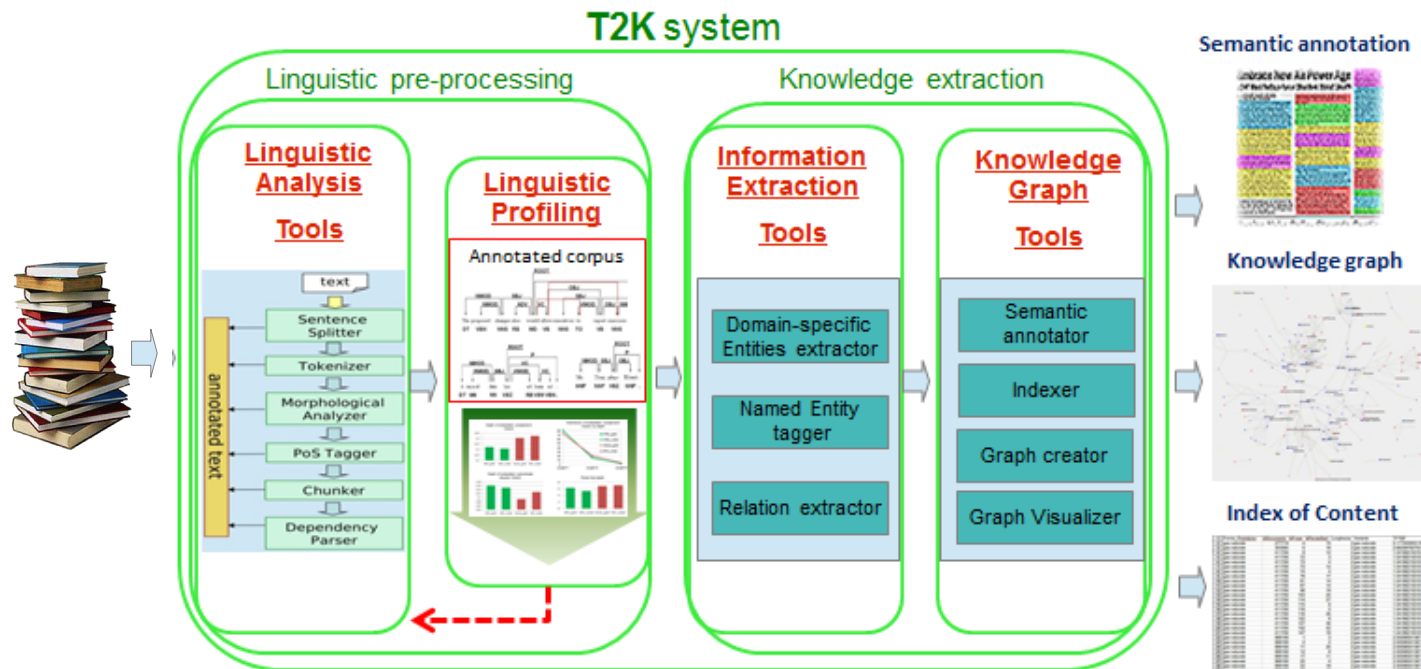
L'accesso alla struttura linguistica

L'accesso al contenuto

# L'accesso al contenuto

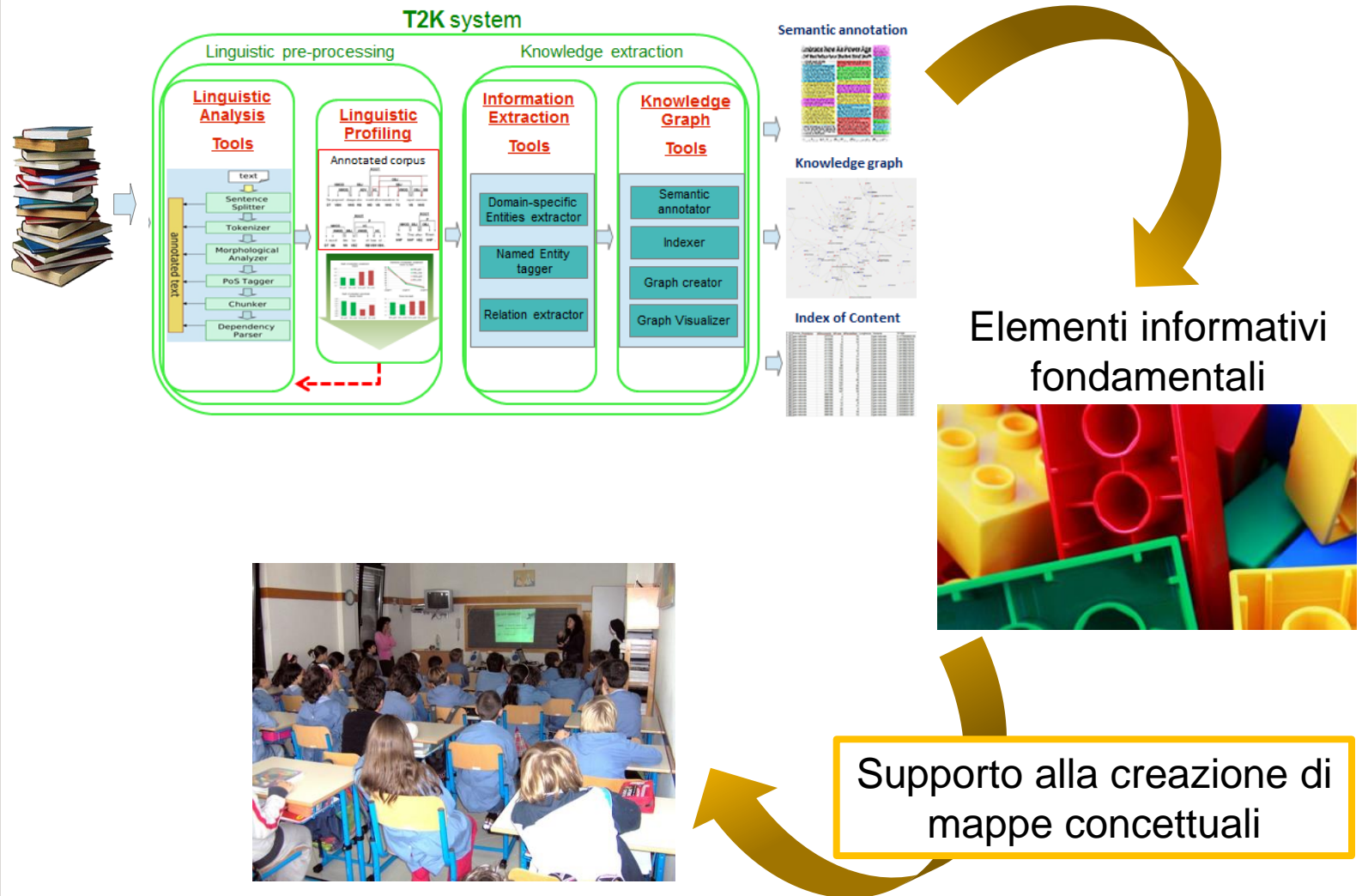


- Text-to-Knowledge (T2K) è una piattaforma che permette di estrarre conoscenza dal testo e organizzarla
  - <http://www.italianlp.it/demo/t2k-text-to-knowledge/>



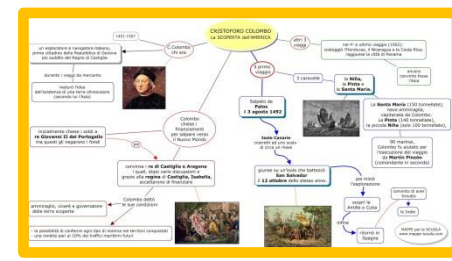
- Usata in diversi contesti applicativi
  - Guide turistiche, brevetti, testi giuridici, cataloghi commerciali, ecc...

# Un ausilio all'organizzazione dei contenuti





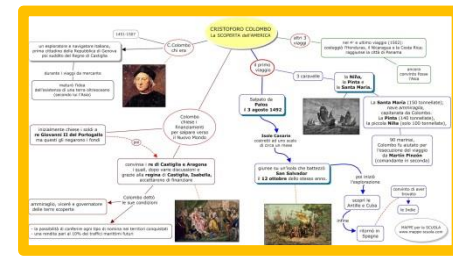
# L'accesso al contenuto



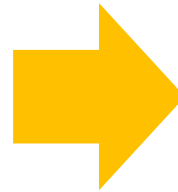
# L'accesso al contenuto



- Termini rilevanti**
- edificio
- affresco
- città
- duomo di Siena
- arte italiana
- colonne
- Giudizio Universale
- storie di San Francesco
- arte classica
- architettura**
- gotico internazionale
- ciclo di affreschi
- pulpito del duomo
- volte a crociera
- tradizione bizantina
- basilica superiore
- ...

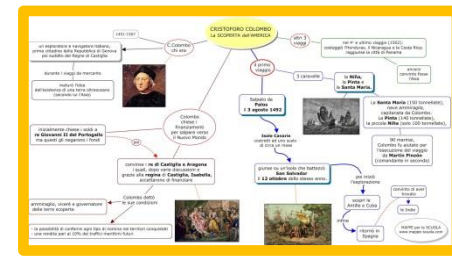


# L'accesso al contenuto



**Termini rilevanti**

- edificio
- affresco
- città
- duomo di Siena
- arte italiana
- colonne
- Giudizio Universale
- storie di San Francesco
- arte classica
- architettura**
- gotico internazionale
- ciclo di affreschi
- pulpito del duomo
- volte a crociera
- tradizione bizantina
- basilica superiore
- ...



## Persone

Giotto  
Cimabue  
Giovanni Pisano  
Simone Martini  
Arnolfo di Cambio  
Ambrogio Lorenzetti  
Cennino Cennini  
Dante  
Duccio di  
Buoninsegna  
...

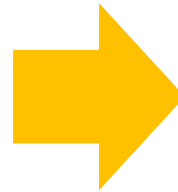
## Luoghi

Roma  
Italia  
Assisi  
Siena  
Firenze  
Pisa  
Padova  
Italia settentrionale  
Saint-Denis  
...

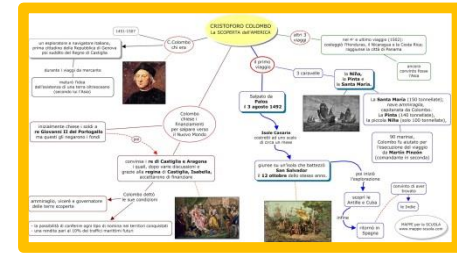
## Organizzazioni

Sacro Romano Impero  
Metropolitan Museum  
Musée de Cluny  
Collezione Salini  
Museo Provinciale  
...

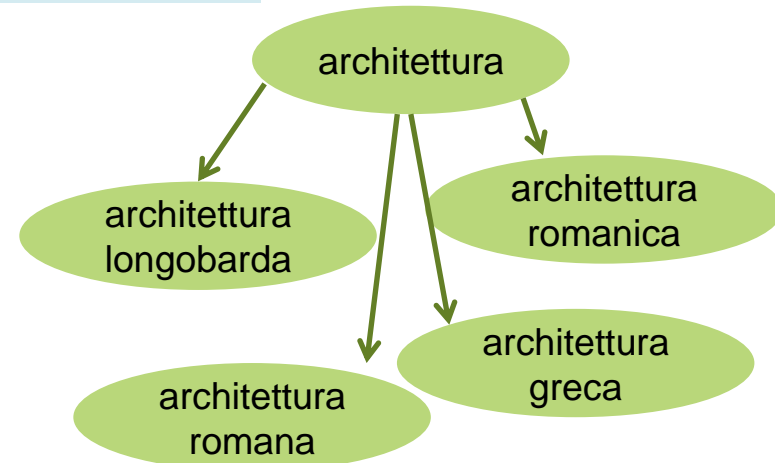
# L'accesso al contenuto



**Termini rilevanti**  
 edificio  
 affresco  
 città  
 duomo di Siena  
 arte italiana  
 colonne  
 Giudizio Universale  
 storie di San Francesco  
 arte classica  
**architettura**  
 gotico internazionale  
 ciclo di affreschi  
 pulpito del duomo  
 volte a crociera  
 tradizione bizantina  
 basilica superiore  
 ...



## Organizzazione tassonomica



### Persone

Giotto  
 Cimabue  
 Giovanni Pisano  
 Simone Martini  
 Arnolfo di Cambio  
 Ambrogio Lorenzetti  
 Cennino Cennini  
 Dante  
 Duccio di Buoninsegna  
 ...

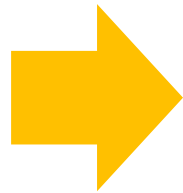
### Luoghi

Roma  
 Italia  
 Assisi  
 Siena  
 Firenze  
 Pisa  
 Padova  
 Italia settentrionale  
 Saint-Denis  
 ...

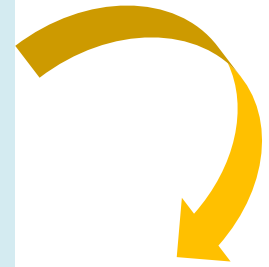
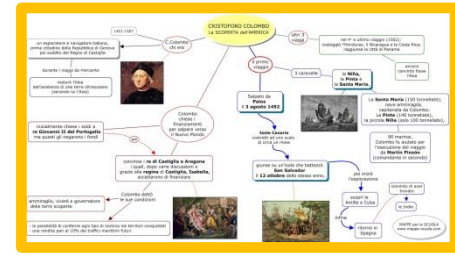
### Organizzazioni

Sacro Romano Impero  
 Metropolitan Museum  
 Musée de Cluny  
 Collezione Salini  
 Museo Provinciale  
 ...

# L'accesso al contenuto



**Termini rilevanti**  
 edificio  
 affresco  
 città  
 duomo di Siena  
 arte italiana  
 colonne  
 Giudizio Universale  
 storie di San Francesco  
 arte classica  
**architettura**  
 gotico internazionale  
 ciclo di affreschi  
 pulpito del duomo  
 volte a crociera  
 tradizione bizantina  
 basilica superiore  
 ...



## Grafo della conoscenza



### Persone

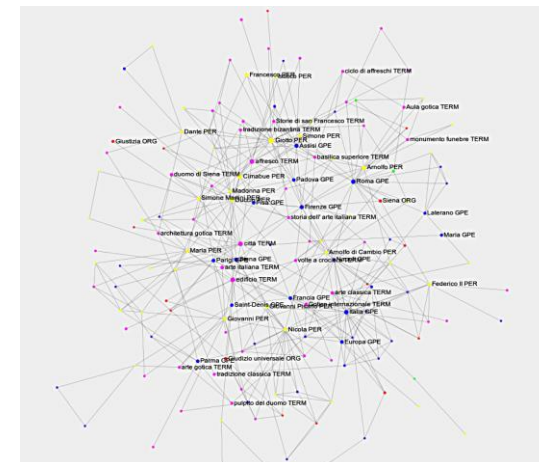
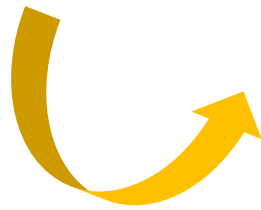
- Giotto
- Cimabue
- Giovanni Pisano
- Simone Martini
- Arnolfo di Cambio
- Ambrogio Lorenzetti
- Cennino Cennini
- Dante
- Duccio di Buoninsegna
- ...

### Luoghi

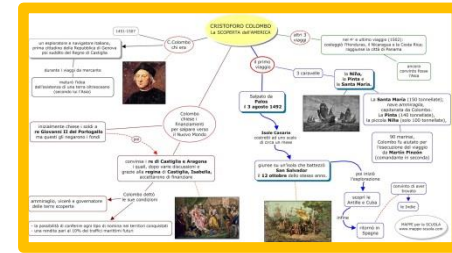
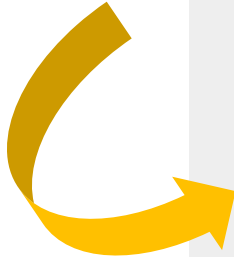
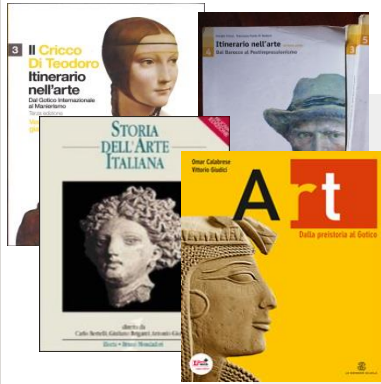
- Roma
- Italia
- Assisi
- Siena
- Firenze
- Pisa
- Padova
- Italia settentrionale
- Saint-Denis
- ...

### Organizzazioni

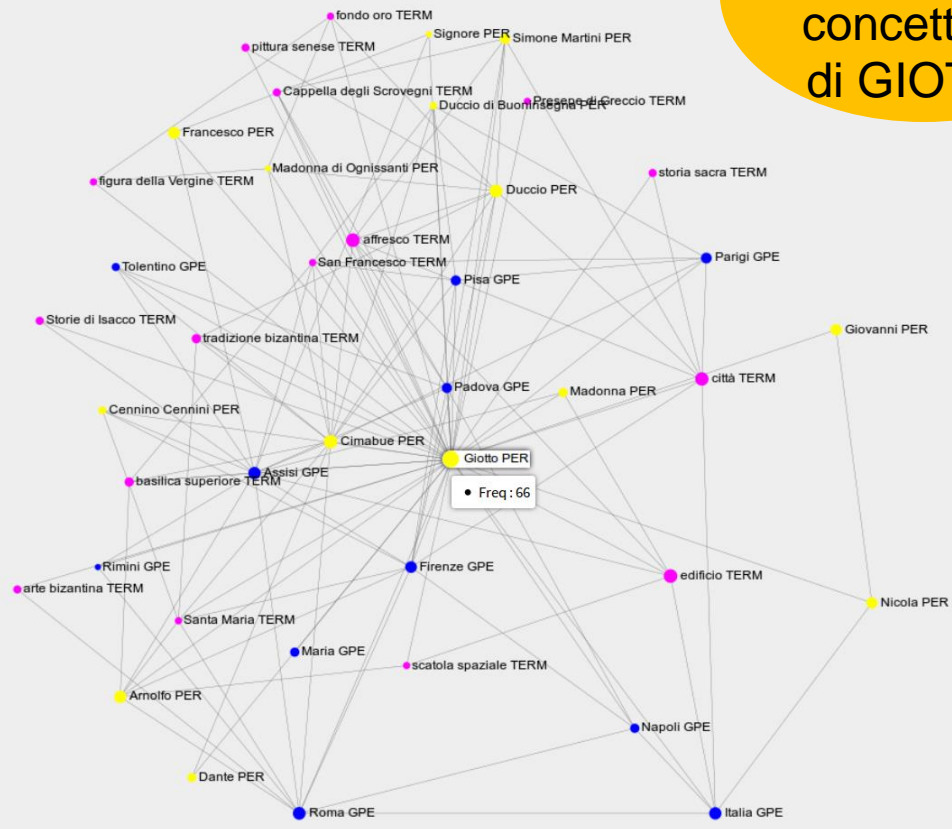
- Sacro Romano Impero
- Metropolitan Museum
- Musée de Cluny
- Collezione Salini
- Museo Provinciale
- ...



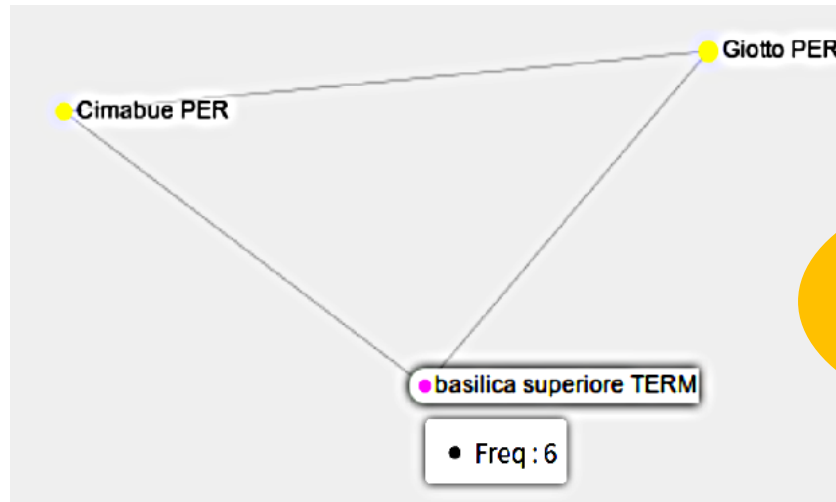
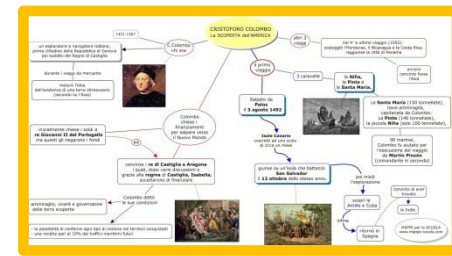
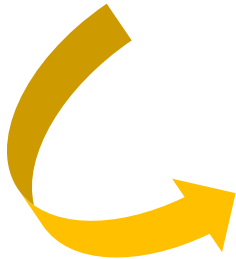
# L'accesso al contenuto



Verso la mappa concettuale di GIOTTO



# L'accesso al contenuto



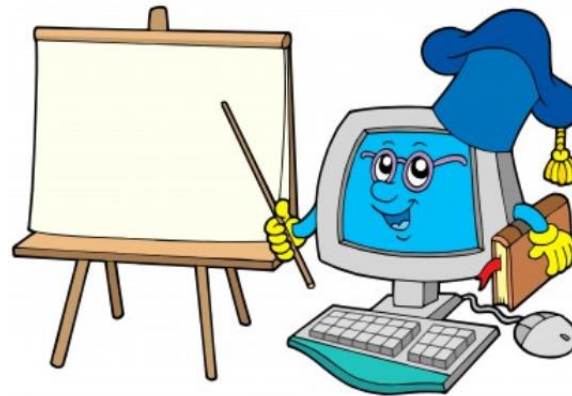
Relazioni  
condivise tra  
due entità

Con queste premesse fondamentali, **Cimabue** affronta negli anni Ottanta la decorazione del capocroce della **basilica superiore** di **San Francesco** ad **Assisi**, dove lavorerà di lì a poco il giovane **Giotto**, impegnato nell'esecuzione delle storie del santo.

# Concludendo

Ricostruzione  
dell'identikit  
linguistico di  
apprendenti

Monitoraggio  
dell'evoluzione  
dell'apprendimento  
linguistico



Supporto alla  
costruzione di  
mappe  
concettuali

Misura della  
distanza  
linguistica tra gli  
apprendenti e i  
materiali  
didattici

Valutazione  
del livello di  
leggibilità  
del testo





# Grazie!

[giulia.venturi@ilc.cnr.it](mailto:giulia.venturi@ilc.cnr.it)

[www.italianlp.it](http://www.italianlp.it)

# Monitoraggio della competenza linguistica di apprendenti L2

- L'interrogativo a questo punto è **se e in che misura** la metodologia proposta possa essere utilmente estesa al monitoraggio della competenza linguistica di apprendenti L2
- Una frase dagli elaborati di studenti L2 della scuola primaria
  - *Io o conosuto subito un mio amico che e diventato il mio amico del cuore con lui o fatto anche moltissimi giochi*
- Diversi ordini di problemi
  - mancanza di punteggiatura
  - errori ortografici

Io o conosuto subito un mio amico che e diventato il mio amico del cuore con lui o fatto anche moltissimi giochi

# Le produzioni di apprendenti L2 e gli strumenti di analisi automatica

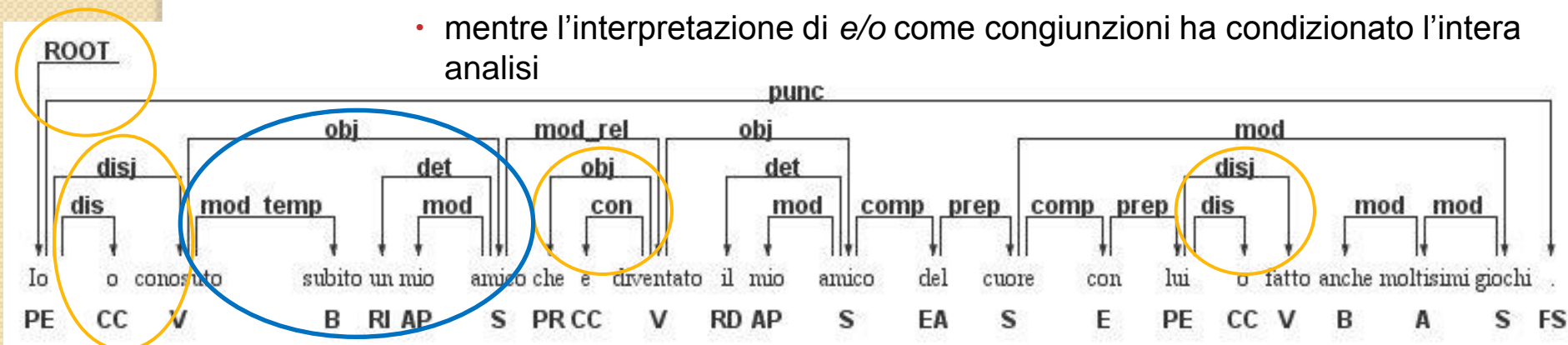
- I diversi ordini di problemi influenzano pesantemente il risultato degli strumenti di analisi automatica
  - analisi morfo-sintattica:
    - **conosuto** per quanto parola sconosciuta interpretato come verbo
    - **e/o** invece di *è/ho* interpretati come congiunzioni

Io o conosuto  
PE CC V

subito un mio amico che e diventato il mio amico del cuore con lui o fatto anche moltissimi giochi .  
B RI AP S PR CC V RD AP S EA S E PE CC V B A S FS

# Le produzioni di apprendenti L2 e gli strumenti di analisi automatica

- I diversi ordini di problemi influenzano pesantemente il risultato degli strumenti di analisi automatica
  - analisi morfo-sintattica
    - **conosuto** per quanto parola sconosciuta interpretato come verbo
    - **e/o** invece di *è/ho* interpretati come congiunzioni
  - analisi sintattica a dipendenze
    - struttura dell'albero fortemente condizionata dagli errori al livello morfo-sintattico
      - *amico* correttamente interpretato come oggetto di *conosuto*
      - mentre l'interpretazione di *e/o* come congiunzioni ha condizionato l'intera analisi



# Che fare?

- Diverse strategie possono essere messe a punto per applicare utilmente la metodologia di monitoraggio anche su produzioni di apprendenti L2
  - recupero degli errori di tipo ortografico
    - caso di **conosuto**: parole sconosciute recuperate attraverso algoritmi di similarità tra stringhe (es. “Levenshtein distance”)
  - personalizzazione degli strumenti di analisi linguistica automatica
    - analisi morfo-sintattica
      - rendere lo strumento in grado di individuare sequenze di categorie grammaticali non canoniche, indipendentemente dall’informazione estratta dal dizionario
        - caso di **elo**: sequenza pron\_pers(PE) cong\_coord(CC) verbo\_part\_pass(V\_pp)
      - ricorso a categorie sottospecificate
- Ciò nonostante, nella situazione attuale ci sono aspetti della competenza linguistica di L2 che possono essere utilmente monitorati, ad es. competenza lessicale