



Dipartimento di Scienze Politiche, Sociali e della Comunicazione
Laboratorio di Linguistica Computazionale
"Maurice Gross"
Università di Salerno

PoliModalCorpus
verso la costruzione del primo corpus
multimodale di dominio politico in italiano

Daniela Trotta

Teresa Albanese

Annibale Elia

LINGUAGGIO POLITICO



Usi politici
della lingua

Tecniche
discorsive e
argomentative

Meccanismi
mediali della
prassi
politica



CORPORA MULTIMODALI

CORPORA MULTIMODALI

1. *text describing pictures*
2. *audio-video recordings with annotations or codings*
3. *audio-video recordings with transcriptions*



A special case of texts describing pictures often used to describe gestures, prosody or other aspects.

It can also be used to capture features of context or various types of semantic-pragmatic information.

(Allwood et al., 2008)

STATO DELL'ARTE

- *Koutsombogera e Papageorgiou (2013)*
- *CorpusB (Bolasco et al., 2006)*
- *CORPS (Guerini et al., 2013)*
- *CompWHoB (Esposito et al., 2015)*



PoliModalCorpus



UNITÀ DI ANALISI



59

interviste

840

minuti

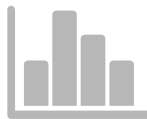
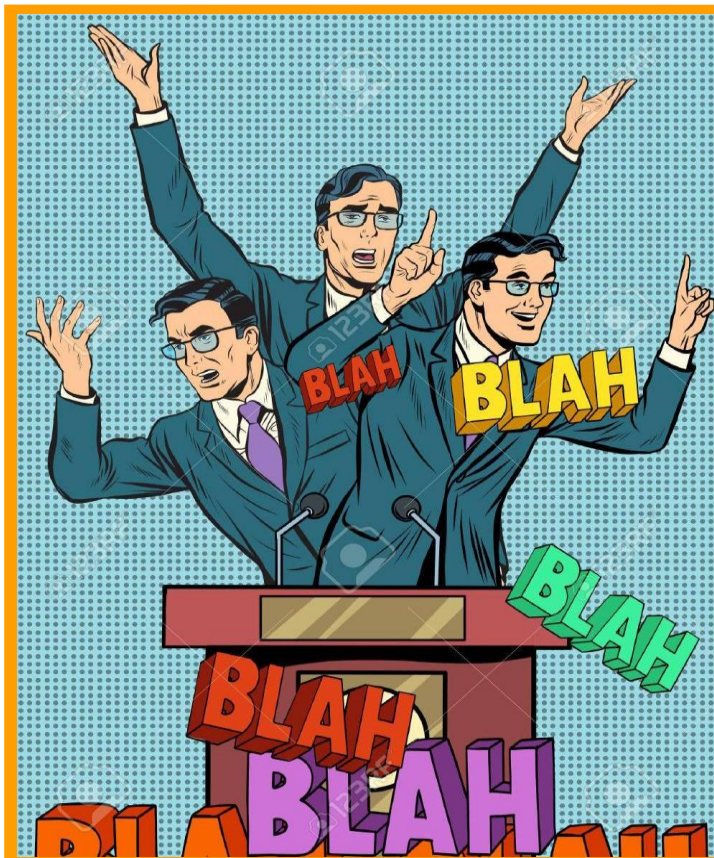
24/09/17 – 14/01/18

arco temporale

RACCOLTA DATI



speech-to-text semi-supervisionato
(Google API + trascrizione manuale)



158543

Token

11231

Type

0.07

TTR



STANDARD DI ANNOTAZIONE

*TEI Guidelines for Electronic Text
Encoding and Interchange*

MARKUP METATESTUALE

<transcriptionDesc ident="HIAT" version="2004"/>

<equipment>

<sourceDesc url="..." user="..." />

<titleStmt>

<recordingStmt>

ESEMPIO

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <!DOCTYPE trascrizioni SYSTEM "unicum.dtd">
3 <trascrizioni>
4   <teiHeader>
5     <transcriptionDesc ident="HIAT" version="2004"/>
6     <equipment>
7       <p>Google API + revisione manuale</p>
8     </equipment>
9     <sourceDesc url="http://www.raipplay.it/video/2017.html" user= "RaiPlay"/>
10    <titleStmt>
11      <title>Mezz'ora in più </title>
12    </titleStmt>
13    <recordingStmt>
14      <recording title="Mezz'ora in più" type="talk show politico" dur="01:01:46"/>
15      <host name="Lucia Annunziata"/>
16      <date>22 ottobre 2017</date>
17      <location>studio</location>
18      <format_type>talk show politico</format_type>
19    </recordingStmt>
20  </teiHeader>
```

MARKUP INTRATESTUALE

<speech id="xxx">

<u who="name and surname" role="host or job of interviewer"
gender="m o f"/>

<pause dur = "...s"/>

<del type="repetition" or "truncation" or "falseStart"/>

<vocal type="semi-lexical" or "non-lexical" desc="..." dur="...s"/>

<u trans="overlap" who="#nome e cognome" role="job" gender="m
o f"/>

ESEMPIO

```
1 <speech id="005">
2 <u who="Lucia Annunziata" role="host" gender="f">Buongiorno Matteo Renzi, Segretario del
  Partito Democratico.</u>
3 <u who="Matteo Renzi" role="PD party secretary" gender="m">Buongiorno!</u>
4 <u who="Lucia Annunziata" role="host" gender="f"><p>Bentornato, è quasi un anno che
  lei<vocal type="semi-lexical" desc="ehm" dur="2s"/>E' una settimana molto importante
  non<del type="falseStart"/>lei è impegnato in un lunghissimo viaggio di 107 tappe. Ne ha
  fatte solo 21, un tour de force.</p></u>
5 <u who="Matteo Renzi" role="PD party secretary" gender="m"><p>Beh solo 21.<del
  type="repetition">La prima, la prima</del>settimana, in treno.<del type="truncation"/>
  </p></u>
6 <u trans="overlap" who="#Lucia Annunziata" role="host" gender="f"><p>Sole su 107.<del
  type="repetition">In senso, in senso</del>, io<del type="falseStart"/>si. E poi insomma
  c'è tutto, ci son tutte le polemiche<vocal type="semi-lexical" desc="ehm"/>il Governo,
  le banche, le solite cose che la seguono diciamo. Io però vorrei farle solo una domanda
  iniziale: io ho visto che in questi mesi passati lei ha maturato nel suo discorso
  pubblico, tra le altre cose, una con<del type="truncation"/>un continuo richiamo
  diretto...
```



PAUSE

REPETITION

DEL

VOCAL

OVERLAP

204

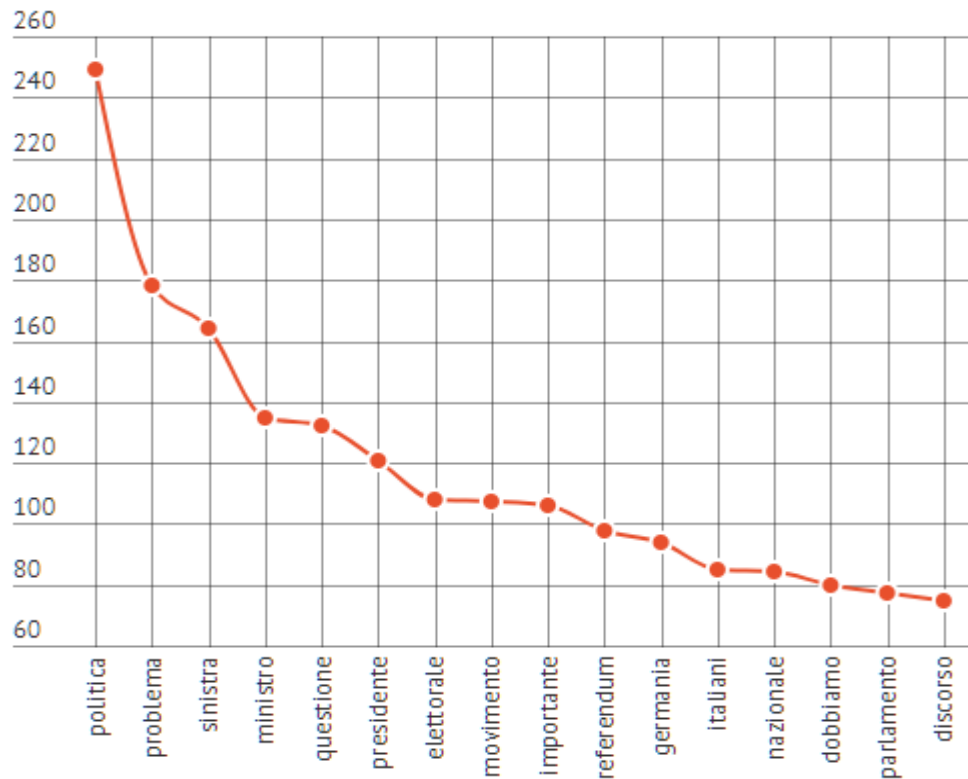
818

465

1346

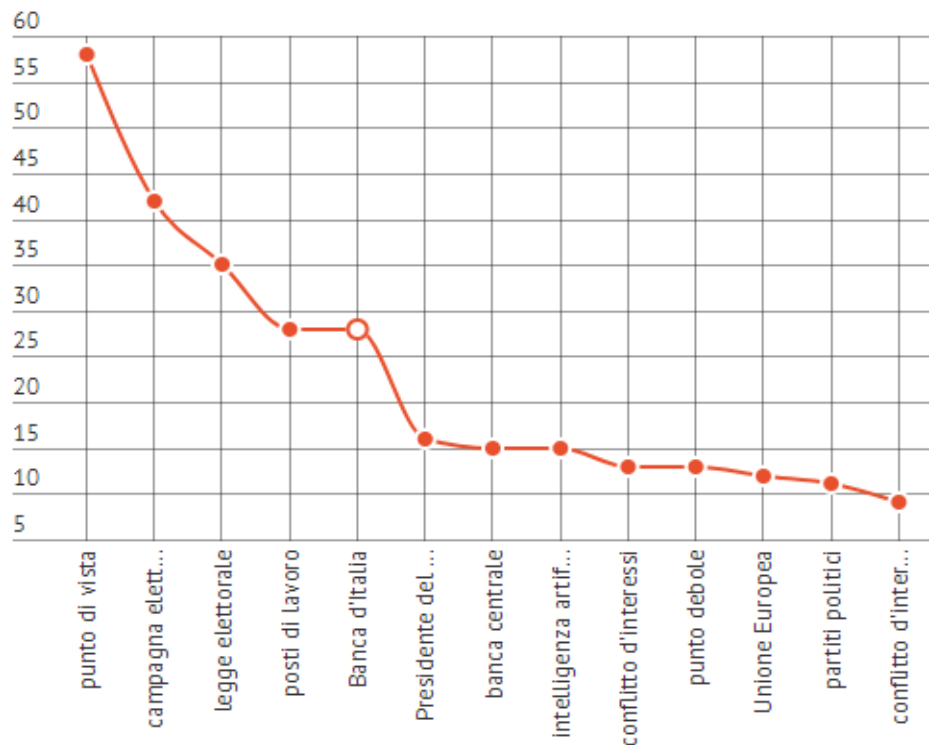
1008

PAROLE SEMPLICI



politica	249
problema	178
sinistra	164
ministro	135
questione	132
presidente	121
elettorale	108
movimento	107
importante	106
referendum	98
germania	94
italiani	85
nazionale	84
dobbiamo	80
parlamento	77
discorso	75

MULTIWORD EXPRESSION



punto di vista	58
campagna elettorale	42
legge elettorale	35
posti di lavoro	28
Banca d'Italia	28
Presidente del Consiglio	16
banca centrale	15
intelligenza artificiale	15
conflitto d'interessi	13
punto debole	13
Unione Europea	12
partiti politici	11
conflitto d'interesse	9

LA LEGGE
ELETTORALE
COSÌ DARÀ UN
VINCITORE CERTO

ODDIO,
CHI?



ETICHETTE TERMINOLOGICHE

1	<ECON>	609
2	<POL>	236
3	<DIR>	216
4	<DIGE>	144
5	<INF>	80
6	<MEST>	37
7	<FIG>	37
8	<ING>	37
9	<MED>	25
10	<MIL>	17



SVILUPPI FUTURI

- ampliamento del corpus
- analisi unità polirematiche
- arricchimento dello schema di annotazione
- accessibilità



gestualità,
mimica facciale

<kinesic>

contenuti

<del

"repetition"

"truncation"

"falseStart"

...



picchi e tensione vocale

<shift>

THANKS



Daniela Trotta d.trotta9@studenti.unisa.it

Teresa Albanese t.albanese@studenti.unisa.it

Annibale Elia elia@unisa.it