

## ApproccioCapacitante® e costituzione del Corpus Anchise: un dataset di 320 conversazioni professionali

Andrea Bolioli CELI - Language Technology, Torino  
Nicola Benvenuti Università degli Studi di Torino (tesista)  
Pietro Vigorelli Gruppo Anchise, Milano

---

---

---

---

---

---

---

---

## Una ricerca nata dall'incontro tra ricercatori clinici e ricercatori computazionali

La cooperazione tra ricercatori clinici e ricercatori computazionali è difficile a causa del differente focus attenzionale:

- Per il clinico al centro dell'attenzione ci sono *le persone*,
- Per il tecnico ci sono *i dati* che si riferiscono alle persone.

In questa ricerca abbiamo trovato un punto d'incontro fecondo nello studio delle *parole* proferite dai pazienti con demenza.

- Per il clinico sono un'espressione della persona,
- per il tecnico sono dati oggettivi studiabili col metodo computazionale.

Il clinico temeva l'effetto dell'appiattimento dei dati che fa perdere il valore delle soggettività.

La ricerca effettuata su un grande numero di dati, invece, ha prodotto risultati simili a quelli che erano stati ottenuti con studi su singoli casi o piccole serie di conversazioni.

---

---

---

---

---

---

---

---

### SCOPO

**Creazione e analisi linguistica** del Corpus Anchise 320, composto da 320 trascrizioni di conversazioni in lingua italiana tra operatori sanitari e pazienti con demenza.

Il Corpus Anchise è costituito da **20.588 turni di dialogo**, di cui 10.193 turni di pazienti e 10.381 turni di operatori sanitari.

222.588 parole (*tokens*)

14.513 parole distinte (*types*)

---

---

---

---

---

---

---

---

Un esempio di frammento di dialogo tra paziente (P) e operatore (O). I dati personali, tra cui i nomi propri, sono stati anonimizzati in accordo al GDPR.

7	P	Eh ma mia figlia... è dura, è dura.
8	O	Lo sa [NOME] che da quando abbiamo iniziato a vederci lei parla molto, molto meglio?
9	P	Ma a casa mia no! Loro non capiscono. Han detto che non capiscono niente...
10	O	Forse sono loro che non capiscono.

4

---

---

---

---

---

---

---

---

---

---

**METODI**

L'analisi automatica dei testi è stata effettuata con la libreria di *Natural Language Processing "StanfordNLP"* per l'italiano e poi corretta manualmente.

Abbiamo effettuato

- la **tokenizzazione** (scomposizione del testo in *token*, cioè parole e segni di punteggiatura),
- la **lemmatizzazione** (annotazione del lemma della parola),
- il **Part of Speech Tagging** (annotazione della parte del discorso, ovvero Nome, Verbo, Articolo, Aggettivo, Pronome, ecc),
- l'annotazione delle **dipendenze sintattiche** (annotazione delle relazioni tra le parole della frase, tra le quali Soggetto, Oggetto, Ausiliare, Apposizione, ecc).

5

---

---

---

---

---

---

---

---

---

---

Esempio di una porzione di frase con le annotazioni linguistiche:

ID	TOKEN	XPOS	LEMMA	FEATS	HEAD	DEPREL
3	che	PRON	che	PronType=Rel	4	nsubj
4	fa	VERB	fare	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin	2	acl:relcl
5	un	DET	uno	Definite=Ind Gender=Masc Number=Sing PronType=Art	6	det
6	po'	ADV	poco		7	advmod
7	fatica	NOUN	fatica	Gender=Fem Number=Sing	4	obj
8	a	ADP	a		9	mark
9	parlare	VERB	parlare	VerbForm=Inf	4	xcomp

6

---

---

---

---

---

---

---

---

---

---



**CONCLUSIONI**

Il Corpus Anchise 320, stratificabile per età, sesso e grado di demenza è ora disponibile per ulteriori ricerche di rilevanza clinica

su

- *diagnosi precoce,*
- *stadiazione delle demenze,*
- *validazione di approcci terapeutico-riabilitativi tra cui in particolare l'Approccio Capacitante di cui riferiamo in una successiva comunicazione in questo stesso congresso.*

---

---

---

---

---

---

---

---