

Quaderni del Dipartimento di Linguistica  
Università della Calabria

n. 31/2023-24

Quaderni del Dipartimento di Linguistica n. 31/2023-24



CACUCCI  EDITORE  
BARI

ISBN 979-12-5965-477-9



9 791259 654779

---

**Quaderni del Dipartimento di Linguistica**  
**Università della Calabria**

---

n. 31/2023-24

## **Direttore**

Roberto Guarasci, UNIVERSITÀ DELLA CALABRIA

## **Comitato scientifico**

Francesco Altimari, UNIVERSITÀ DELLA CALABRIA

Mario Caligiuri, UNIVERSITÀ DELLA CALABRIA

Carlo Spartaco Capogreco, UNIVERSITÀ DELLA CALABRIA

Mario Caterini, UNIVERSITÀ DELLA CALABRIA

Angela Costabile, UNIVERSITÀ DELLA CALABRIA

Belinda Crawford, UNIVERSITÀ DELLA CALABRIA

Ines Crispini, UNIVERSITÀ DELLA CALABRIA

Maria Mirabelli, UNIVERSITÀ DELLA CALABRIA

Luciano Romito, UNIVERSITÀ DELLA CALABRIA

Anna Rovella, UNIVERSITÀ DELLA CALABRIA

Giuseppe Spadafora, UNIVERSITÀ DELLA CALABRIA

## **Responsabile di redazione**

Assunta Caruso, UNIVERSITÀ DELLA CALABRIA

## Quaderni del Dipartimento di Linguistica Università della Calabria

La rivista “Quaderni del Dipartimento di Linguistica” è una rivista scientifica inerente gli ambiti di ricerca del Dipartimento di Culture, Educazione e Società ed è censita dall’ANVUR per l’Area 10 – Scienze dell’antichità, filologico-letterarie e storico-artistiche e l’Area 11 – Scienze storiche, filosofiche, pedagogiche e psicologiche.

Lo scopo della rivista è di creare un’occasione in cui studiosi, docenti e ricercatori possano condividere idee sulla ricerca e didattica svolta nel campo di studio proposto, attraverso uno sguardo collaborativo e innovativo. La rivista include diverse sotto-aree con l’intento di incoraggiare la condivisione di una più ampia gamma di esperienze, risultati, scambi di idee.

I contributi potranno essere presentati in diverse lingue allo scopo di rafforzare l’aspetto multilingue della rivista ed evidenziare la pluralità di culture e stili comunicativi con i quali la comunità universitaria entra in contatto durante la quotidianità accademica.



*Quaderni del Dipartimento di Linguistica*

UNIVERSITÀ DELLA CALABRIA

31/2023-24

**Echi di cittadinanza. Ricerca, esperienze e  
contaminazioni tra Calabria e Argentina.**

*a cura di*

Silvia Morelli

Anna Rovella

Mario Francisco Benvenuto

Rossella Michienzi

CACUCCI  EDITORE  
BARI

---

PROPRIETÀ LETTERARIA RISERVATA

---

© 2024 Cacucci Editore – Bari

Via Nicolai, 39 – 70122 Bari – Tel. 080/5214220

<http://www.cacuccieditore.it> e-mail: [info@cacucci.it](mailto:info@cacucci.it)

Ai sensi della legge sui diritti d'Autore e del codice civile è vietata la riproduzione di questo libro o di parte di esso con qualsiasi mezzo, elettronico, meccanico, per mezzo di fotocopie, microfilms, registrazioni o altro, senza il consenso dell'autore e dell'editore.

## Sommario

Introduzione	9
<i>Silvia Morelli</i> , Premisa	11
<i>Maria Giuseppina Bartolo</i> , Interventi <i>Evidence-Based</i> per prevenire e contrastare il bullismo e il cyberbullismo	13
<i>Francesco Craig</i> , Autolesionismo non suicidario come manifestazioni del malessere psicologico adolescenziale	33
<i>Anna Lisa Palermi</i> , La fenomenologia del bullismo e del cyberbullismo a scuola	47
<i>Natalia Forlini</i> , <i>Fernando Fariás</i> , El impacto subjetivo del trabajo grupal en el ámbito educativo	65
<i>Verónica Torres</i> , <i>Guillermo Dezorzi</i> , La adolescencia desde una concepción del sujeto	75
<i>Mario F. Benvenuto</i> , <i>Rossella Michienzi</i> , Escenarios inéditos de enseñanza-aprendizaje en la didáctica de Lenguas Extranjeras	95
<i>Donata Chiricò</i> , L'educazione democratica alla prova della sordità	119
<i>Paula Marini</i> , Franco Carbonari, Temporalidades, vínculos y lo escolar en el presente	133
<i>Lucia Montesanti</i> , Francesca Veltri, Politiche di protezione e processi di socializzazione dei minori non accompagnati in Italia, tra impegno civile e prospettive giuridiche	157
<i>Rocco Servidio</i> , <i>Martina Basilico</i> , Il gioco online problematico: uno studio esplorativo	183
<i>María Isabel Pozzo</i> , Prolegómenos para el diálogo intercultural y la comunicación científica en educación	227
<i>Giovanna Vingelli</i> , L'educazione alla sessualità	253
<i>Paula Caldo</i> , ¿Dónde están las mujeres en los archivos?	277
<i>Martin Critelli</i> , L'applicazione di reti neurali convoluzionali per l'estrazione automatica della conoscenza dei beni culturali da archivi multimediali	297
<i>Adriana Hereñú</i> , Socialización de experiencias de indagación en fuentes documentales y testimonios como enfoque metodológico en investigación educativa	309
<i>Claudia Lanza</i> , Dal riconoscimento della scrittura manuale alla categorizzazione automatica dei dati	331
<i>Anna Rovella</i> , Resilienti o persistenti?	343





# Dal riconoscimento della scrittura manuale alla categorizzazione automatica dei dati

Claudia Lanza\*

**Abstract:** This paper presents the work carried out for the digitization and transcription of a sample of clinical records belonging to the Neurogenetics Center archive in Lamezia Terme (CZ) dealing with the treatment of neurodegenerative illness. The main purposes of this work are equally that of preserving the cultural value of scientific research on the disease and of creating a digital categorized infrastructure within which the informative tissue, resulted by the handwritten text recognition and transcription, can be exploited for computational studies and, thus, facilitating the retrieval operations over clinical data to use in a wide-scope health system research.

**Keywords:** Digitization; Transcription; Semantic Annotation; Categorization; Neurodegenerative Studies.

## 1. Introduzione

Il lavoro di digitalizzazione, oggetto del presente contributo<sup>1</sup>, ha riguardato la documentazione dell'archivio del Centro Regionale di Neurogenetica (CRN) sito a Lamezia Terme (CZ) e costituito dalle cartelle cliniche dei pazienti affetti da malattie neurodegenerative. L'archivio a scaffalatura si compone di 12,860 cartelle cliniche di pazienti, che datano dal 1992. Il CRN è ubicato all'interno dell'Azienda Sanitaria Provinciale di Catanzaro con la mission di sviluppare attività clinico-assistenziali e ricerche scientifiche sullo studio delle Malattie Neurodegenerative, con un particolare interesse per la Malattia di Alzheimer, le demenze e le malattie neurologi-

---

\* Dipartimento di Culture, Educazione e Società, Università della Calabria. claudia.lanza@unical.it.

<sup>1</sup> Per il presente contributo si descriveranno le diverse fasi di attività di ricerca portate avanti in prima persona, nello specifico il lavoro di riconoscimento della scrittura manuale a partire dalla documentazione digitalizzata del Centro e la categorizzazione automatica delle informazioni.

che rare a carattere ereditario. Tra gli obiettivi principali perseguiti durante attività di ricerca, qui presentata, vi è anzitutto la possibilità di garantire la conservazione e la tutela dei documenti clinici per le finalità di studio e ricerca. Poter usufruire nel tempo di documenti digitalizzati e strutturati sullo studio delle malattie neurodegenerative consentirebbe, infatti, di costruire percorsi di ricerca scientifica basati su uno studio diacronico in grado di contribuire ad indagare gli sviluppi nella genealogia per la ricostruzione degli indici di familiarità nelle malattie neurodegenerative.

La digitalizzazione delle cartelle cliniche presenta ulteriori vantaggi in termini di utilizzo delle informazioni da una prospettiva computazionale. Il CRN dispone di un proprio sistema di registrazione dei dati relativi ad ogni paziente attualmente o precedentemente in cura presso la struttura. Tale sistema è strutturato in macroaree nelle quali sono inserite le informazioni relative, ad esempio, ai dati anagrafici, ai dati antropometrici, ai dati anamnestici, al diario clinico, ai test cognitivo-comportamentali, ai test neuropsicologici, alle indagini strumentali, agli esami di laboratorio, agli esami genetici, agli alberi genealogici (realizzati attraverso l'utilizzo di applicativi software in house), al sistema di prenotazioni, alle esenzioni, al piano terapeutico e, se presenti, alle prestazioni esterne. Pertanto, la digitalizzazione in formato strutturato delle informazioni contenute all'interno della documentazione clinica, attualmente prodotta in formato cartaceo, faciliterebbe i processi di analisi e recupero automatico delle informazioni<sup>2</sup>.

Inoltre, la digitalizzazione della documentazione permette anche la trasposizione in forma processabile di testi redatti con scrittura manuale, con il conseguente vantaggio di poter esercitare un controllo semantico sui contenuti agevolando i processi di ricerca delle informazioni e gestione della conoscenza. Nel caso specifico, ad esempio, la registrazione delle informazioni in formato digitale di quanto scritto manualmente dai neurologi curanti porta ad una accurata gestione della variabilità semantica per documenti che contengono una narrazione della storia clinica del paziente, mediante l'implementazione di modelli interpretativi automatici per la categorizzazione delle informazioni.

---

<sup>2</sup> Per un approfondimento sul tema degli archivi nell'era digitale si rimanda al testo di Valacchi, F. 2006. *La memoria integrata nell'era digitale. Continuità archivistica e innovazione tecnologica*. Corazzano: Titivillus Edizioni.

## 2. La metodologia di digitalizzazione

La documentazione contenuta all'interno dell'archivio della struttura del CRN segue l'attività clinico-assistenziale del Centro ripercorsa attraverso la sedimentazione di documenti all'interno delle cartelle cliniche di ogni paziente, e che inquadrano le molteplici fasi di trattamento della malattia neurodegenerativa da parte della struttura. Il lavoro di digitalizzazione ha interessato documenti originali forniti dal soggetto produttore.

L'acquisizione della documentazione clinica ai fini dell'attività di ricerca di digitalizzazione e trascrizione ha osservato i seguenti passaggi:

- a. Delimitazione dell'ambito di studio.
- b. Definizione dei criteri di selezione dei documenti in base agli scopi di ricerca.
- c. Individuazione dei documenti.
- d. Analisi delle tipologie documentali.
- e. Estrazione dei dati per gli scopi di ricerca.

Nel dettaglio, la delimitazione dell'ambito di studio ha riguardato la selezione delle cartelle cliniche da sottoporre alle fasi di digitalizzazione e trascrizione. Le cartelle cliniche prese in considerazione hanno interessato soltanto i pazienti deceduti affetti da malattia di Alzheimer (AD), pari a circa 800 cartelle. All'interno di questo campione di riferimento è stato individuato un altro set di dati per avviare le fasi di addestramento del modello di riconoscimento della scrittura manuale. Tale raggruppamento, pari al 10% del campione di riferimento, è stato supportato dall'affiancamento con esperti del dominio, interni al soggetto produttore, che ne hanno individuato le sezioni più rilevanti e pertinenti agli scopi di ricerca (Hassanzadeh et al. 2016). La successiva fase ha riguardato l'analisi delle tipologie documentali presenti all'interno delle cartelle prese in considerazione, in particolare, è stata esaminata la distribuzione dei dati in relazione alle seguenti sezioni individuate con la terminologia usata dal soggetto produttore non puntuale sotto il profilo documentale ma maggiormente esplicita per il dominio medico: anamnesi patologica prossima, anamnesi patologica remota, visite di follow-up, esami di laboratorio, esami psicologici, test strumentali, storia clinica del paziente. Le sezioni riferite all'anamnesi hanno rappresentato quelle più significative ai fini dell'avvio del processo di trascrizione del testo dal momento che contengono una descrizione accurata della malattia da parte del medico attraverso la scrittura manuale, così come mostrato nelle Figure 1, 2, 3. L'obiettivo di questa

attività di ricerca è strettamente collegato al processo di riconoscimento del testo e della sua trascrizione ai fini di un utilizzo delle informazioni testuali sottoforma di linguaggio normalizzato e in formato processabile che supporti l'inserimento dei dati clinici all'interno del database in maniera automatica e precisa.

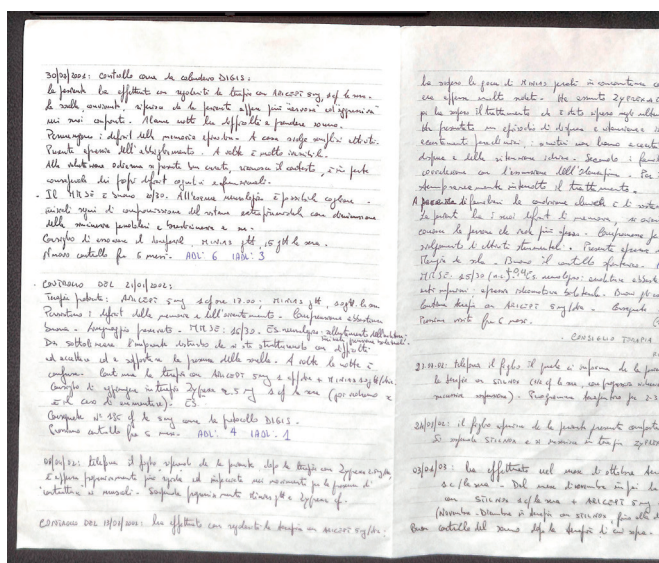


Figura 1. Descrizione anamnestica redazione manuale, documento digitalizzato.

Annuncio il Kalligrafia a 10/3

27/8/95 Luchello: ... il paziente è un uomo sposato. Ha cinque  
 la moglie un bel lavoro e gli altri di vivere al Tolosa.  
 La signora per lui il Kalligrafia e dice che Tolosa fu a 8/10  
 questa sommità. Infortunio anche essere accorto e non  
 e se non è associato. L'unico a dire per il tavolo  
 e alla moglie. Ha avuto delle intossiche e  
 sempre nelle urine due volte è stato ricoverato a e e ma  
 per lo hanno detto di andare per l'operazione.  
 Completamente agitato. L'unico di Kalligrafia  
 Grave amnesia. Ipotonia. Ipotonia. Ha ipotonia  
 con un po' di ipotonia a 10/3.  
 Segno di Kalligrafia +++  
 Apertosi nel cammiera, non si vede nel Kalligrafia  
 Adoro per l'operazione e molto e di notte da  
 dormire. Ricoverato. Tolosa 3/10 + 3.  
 Atene amnesia totale

11/1/97  
 Luchello amnesia. Il paziente ha avuto un lungo periodo di  
 agitazione. A 10/3 di Kalligrafia con Sberzo e Kalligrafia  
 e lui ebbe un grosso crollo. Tolosa si è poi esultato.  
 Ricoverato da bene, mangia molto e si è boccato per 2  
 anni. Non cammina più. Ricoverato un'altra  
 A 10/3 di Kalligrafia. A 10/3 di Kalligrafia.

Le micelose sono cessate.  
 Sberzo spaccatura. Ipotonia emilato oblio  
 ed è 10/3 di Kalligrafia.  
 Facies amnesia. Palmo. Mentone. Sberzo. grigio  
 blut e grigio. Kalligrafia.  
 Continua da moglie a dare l'Kalligrafia una cosa  
 Sberzo!

30/12/2011 Vago a sapere che il paziente è oblioso. In novembre 2011  
 a Vago dove per aiuto a vivere della Togli e dove ha un'auto  
 e a 10/3 di Kalligrafia. La comunicazione fa  
 da Vago (nipote del pz) e mio di Kalligrafia. Case 5/10/12

Figura 2. Descrizione anamnestica redazione manuale, documento digitalizzato.

Adesso da un punto di vista comportamentale è  
 piuttosto inquieto  
 Tappa 10/15 (aspetta)  
 Tappa 25 Relazione  
 Disturbi di comportamento da 8 anni circa con  
 progressivi problemi della memorizzazione. Attualmente  
 disorientata. T.S. Approssia dell'abbigliamento  
 da qualche tempo particolarmente inquieto  
 linguaggio inteso. Non presenta al momento agnosie  
 per i visi. Incontinenza urinaria e fecale. La  
 copia dei disegni mostra il fenomeno del  
 closing-in - operabilità comportamentale  
 L'evoluzione e i segni della paziente dipendono  
 per volume tipo Alzheimer

Figura 3. Descrizione anamnestica redazione manuale, documento digitalizzato.

I dati clinici presenti nelle cartelle analogiche sono stati resi disponibili in forma digitale attraverso un processo di riconoscimento della grafia manuale, utilizzando un software opportunamente addestrato con un training set estratto dal corpus di documenti presi a riferimento. Quest'ultimo è stato costituito da una parte delle sezioni anamnestiche delle cartelle cliniche che sono state trascritte attraverso l'uso del software Transkribus (Nockels 2022) per poter riconoscere non solo i caratteri, ma anche le diverse tipologie di redazione manuale. L'addestramento del modello per il riconoscimento della scrittura manuale ai fini dell'implementazione di una trascrizione automatica ha previsto un'associazione di tipo supervisionata per ogni porzione di testo. I testi trascritti risultanti sono stati poi utilizzati per le fasi di automatizzazione del riconoscimento della scrittura, per l'annotazione semantica e per le successive fasi di categorizzazione dei dati nel database del Centro. Come mostrato in Figura 4, le sezioni che sono state processate in fase di addestramento attraverso Transkribus hanno riguardato le anamnesi patologiche prossime e remote dei pazienti in cura presso il CRN inserite all'interno della documentazione clinica digitalizzata, caratterizzate da un consistente utilizzo della grafia per la descrizione narrativa della sintomatologia. La trascrizione di tale documentazione ha dunque permesso di ottenere in output una serie di testi in formato processabile per le seguenti operazioni di annotazione semantica. Come sottolineato

da Bozzi (2000), «Poiché i documenti consistono principalmente di testo scritto o stampato, la maggior parte dell'informazione può essere ottenuta dall'analisi di tale testo; per questo motivo un passo fondamentale è la trascrizione automatica della sequenza dei caratteri di testo in stringhe di caratteri digitali processabili automaticamente da un computer».

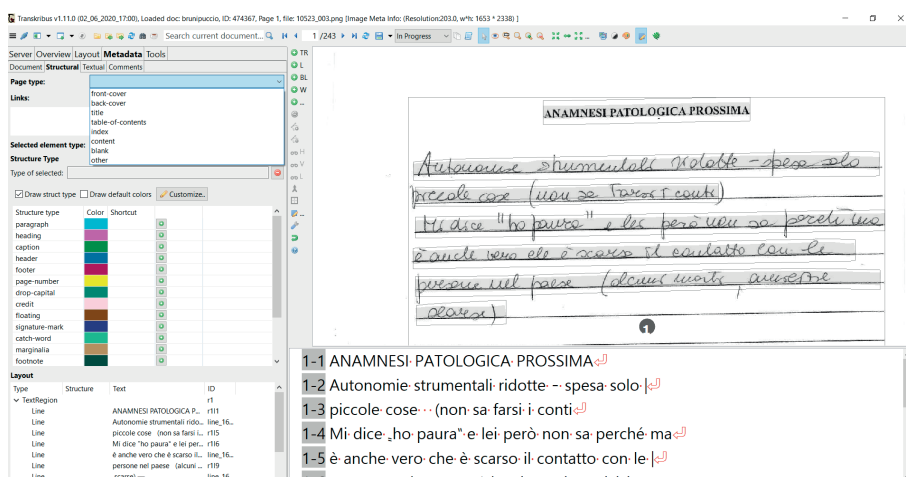


Figura 4. Piattaforma di trascrizione su Transkribus.

### 3. Categorizzazione dei dati per la realizzazione di una base di conoscenza

Il corpus di trascrizioni ottenute dalla fase di riconoscimento della scrittura manuale della sezione anamnestica appartenente alle cartelle cliniche è stato annotato semanticamente<sup>3</sup>. La prospettiva, a tal riguardo, è stata rivolta alla strutturazione delle informazioni testuali all'interno di tali se-

<sup>3</sup> Per i lavori riferiti ai processi di annotazione semantica su documentazione clinica si vedano i lavori di i) Abbas, Asim, Muhammad Afzal, Jamil Hussain, Taqdir Ali, Hafiz S.M. Bilal, Sungyoung Lee, and Seokhee Jeon. 2021. "Clinical Concept Extraction with Lexical Semantics to Support Automatic Annotation" International Journal of Environmental Research and Public Health 18, no. 20: 10564. <https://doi.org/10.3390/ijerph182010564>; ii) Attardi, Giuseppe, Vittoria Cozza, and Daniele Sartiano. 2015. "Annotation and Extraction of Relations from Italian Medical Records." In Proceedings of 6th Italian Information Retrieval Workshop, Cagliari: CEUR Workshop Proceedings; iii) Pinal, Patel, Disha Davey, Vishal Panchal, and Parth Pathak. 2018. "Annotation of a large clinical entity corpus." In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2033-2042.



zioni per poter estrarre una serie di espressioni formali da sintetizzare nelle macro aree del database del Centro, ovvero gli elementi informativi relativi ai dati clinici, farmacologici e cognitivo-comportamentali dei pazienti affetti da malattia neurodegenerativa. La categorizzazione dei dati si basa sull'assegnazione di categorie predefinite a documenti sulla base del loro contenuto e implica un'analisi preliminare del linguaggio naturale, così come puntualizzato da How (2004): «Text categorization (TC) involves the assignment of categories to natural language documents based on the assessment of their contents» (How, 2004). Pawar (2012) sottolinea l'importanza di disporre di un gruppo di categorie per l'avvio del processo: «The task to classify a given data instance into a pre-specified set of categories is known as “text categorization” (TC). Given a set of categories (subjects, topics) and a collection of text documents, it is the process of finding the correct topic (or topics) for each document. »

L'attività di ricerca è stata basata sul potenziamento di categorie preesistenti determinate dal CRN all'interno del proprio database di riferimento attraverso un processo di digitalizzazione e trascrizione della documentazione. Mediante i testi trascritti è stato possibile implementare una serie di regole di annotazione semantica per poter associare a ogni espressione formale, individuata con il supporto degli esperti di dominio e ritenuta rilevante ai fini di un preciso inserimento delle informazioni nelle macro-aree del database, un tag semantico. L'obiettivo è quello di realizzare una base di conoscenza allineata tra la documentazione clinica cartacea e le informazioni raccolte e memorizzate in digitale nel corso dei decenni di attività del Centro. Su questi dati sarà quindi possibile eseguire processi di analisi ed estrazione di modelli di conoscenza descrittivi e predittivi per supportare gli operatori del settore nella gestione e monitoraggio delle demenze.

La Tabella 1 mostra nel dettaglio il gruppo di categorie predeterminate a cui sono state ricollegate diverse espressioni formali rintracciate all'interno delle sezioni anamnestiche della documentazione clinica digitalizzata e trascritta, ritenute rilevanti da parte dei profili di esperti del dominio di riferimento.

<b>ORIENTAMENTO</b> orientamento familiare orientamento ambientale orientamento spaziale	<b>DISTURBI DELLA MEMORIA</b> difficoltà afasia sensoriale memorie vocali riduzioni iniziali verbali assenza linguaggio parafasia semantica parafasia fonologiche pallidita disartria ecclasia afasia no minimum incolata parole	<b>DISTURBI DEL LINGUAGGIO</b> difficoltà afasia sensoriale memorie vocali riduzioni iniziali verbali assenza linguaggio parafasia semantica parafasia fonologiche pallidita disartria ecclasia afasia no minimum incolata parole	<b>AGNOSE</b> segno specchio prosopagnosia disturbi riconoscimento colori	<b>PRASSIE</b> aprassia costruttiva aprassia ideomotoria aprassia ideativa aprassia abbigliamento	<b>PLANNING</b> difficoltà programmazione difficoltà attivazione difficoltà attenzione	<b>GRAFIA CALCOLO LETTURA</b> grafia digradita lettura calcolo	<b>COMPORTAMENTO</b> perseverazione iperattività ricerca igiene personale incapacità critica giudizio impulsività durabilità disinibizione ossessive insight isolamento sociale perdita consapevolezza personale impulsività rigidità mentale inflessibilità rallentamento psicomotorio riduzione iniziale comportamento	<b>DISTURBI AFF</b> ipochondria anaffettività ansia chiusura relaz depressione delirio apatia idee ossessive
---	--	---	--	---	---	--	--	--

Tabella 1. Set di categorie del database CRN.

L'estrazione delle informazioni e l'inserimento dei dati nelle relative categorie sono stati eseguiti attraverso due metodologie: a) l'individuazione e implementazione di regole sintattiche, o *pattern*, e b) la costruzione di reti neurali<sup>4</sup>. Per quanto riguarda la prima metodologia, è stata preliminarmente eseguita un'operazione di marcatura morfosintattica sul corpus dei testi trascritti in modo da poter individuare ogni elemento costitutivo delle frasi in forma automatica con l'applicazione delle regole sintattiche (Mykowiecka 2009). Ogni espressione identificata con il supporto degli esperti di dominio rappresentativa delle categorie sulla diagnosi di Alzheimer è stata mappata con la sua analisi morfosintattica ed estratta dai testi trascritti delle cartelle cliniche (si veda Fig. 5). Questo processo risulta puntuale e preciso, ma non velocizza i processi di riconoscimento dei tag semantici all'interno dei testi perché richiede un livello di specificazione elevato della variabilità semantica distribuita all'interno della documentazione di riferimento su cui avviare le annotazioni semantiche di addestramento. Proprio per tali ragioni la scelta della realizzazione delle reti neurali ha consentito di snellire le fasi di addestramento del modello sui testi trascritti e di ottenere in output una serie di associazioni categoriali più rapide. La successiva Figura 6 mostra come l'interrogazione sui testi risultanti dalla trascrizione su cui è stata eseguita l'operazione di annotazione semantica può portare l'utente a visualizzare la struttura informativa

<sup>4</sup> Negli ultimi anni, le reti neurali hanno contribuito a garantire progressi significativi nel campo delle Named Entities Recognition (NER) e nell'estrazione di concetti medici eliminando la necessità dell'ingegneria delle funzionalità. Le reti neurali ricorrenti (RNN) possono tenere traccia della struttura della frase e delle varie dipendenze e consentire alle informazioni di essere persistenti sulla rete. Si rimanda al lavoro scientifico Pasceri, Erika, Mérième Bouhandi, Claudia Lanza, Anna Perri, Valentina Laganà, Raffaele Maletta, Raffaele Di Lorenzo, and Amalia C. Bruni. 2023. "Neurodegenerative Clinical Records Analyzer: Detection of Recurrent Patterns Within Clinical Records towards the Identification of Typical Signs of Neurodegenerative Disease History". JLIIS.It 14 (2):20-38. <https://doi.org/10.36253/jlis.it-522>, in cui sono forniti tutti i dettagli circa i risultati ottenuti applicando tale metodologia sul corpus delle trascrizioni delle sezioni anamnestiche della documentazione appartenente al CRN.

categorizzata con i tag semantici contenenti dati acquisiti attraverso l'implementazione dei *pattern*.

<b>#PRAXIES</b>
#constructional_apraxia
#ideomotor_apraxia
1) <i>aprassia deambulatoria</i> > NOUN ( <i>aprassia e</i> ) + ADJ ( <i>deambulatoria e</i> )
#ideational_apraxia
#wearing_apraxia

Figura 5. Addestramento per regole sintattiche.

Visite di Follow - up - LIZIDRA 5 mg DRUGS\_VAL , 1 cp ore 900 06/06/2019 : Condizione metabolica generale -  
 EZEMATIS 10 1 DRUGS\_VAL cp + 1/2 cp/die OK. Azotemia 43 DRUGS\_VAL - VIVIFAST (20g/mese). PA : OK. Riesce a  
 svolgere le attività di Vit D base. Limitazioni nelle attività strumentali. Fa colazione da sola . Seguita per l'assunzione della  
 terapia THERAPY\_PRES\_DRUGS . Dimenticanze nel quotidiano non è ben orientata nel tempo Riconosce lo spazio . Capacità  
 ad elaborare i messaggi verbali. Curata nell'aspetto Tranquilla. Consapevole dei deficit . Un pò svogliata. Continua  
 EZEMATIS 10 1 cp THERAPY\_PRES\_DRUGS + 1/2 cp al mattino LIZIDRA 5 DRUGS\_VAL g 1 cp/ ore 1400 VIVIFAST cicli . +  
 terapia di base. Controllo fra sei mesi THERAPY\_FUT - EBIXA 10 DRUGS\_VAL , 1 cp/ e 1/2 al mattino 10 DIC. 2019  
 LIZIDRA 5 mg DRUGS\_VAL 1 cp ore 1400 Vive in compagnia di una badante e (DONEPEZIL) - VIVIFAST - / cicli \_ con una  
 figlia ora venuta a casa. Poco orientata nel tempo , riconosce lo spazio

Figura 6. Struttura informativa categorizzata.

#### 4. Conclusioni

Il lavoro svolto sulla documentazione clinica ha permesso di analizzare il linguaggio medico e strutturarne in termini di un linguaggio formalizzato ai fini di un potenziamento della rete informativa presente nel database del CRN. L'analisi di documenti clinici è stata eseguita sui testi digitalizzati scritti in linguaggio naturale in cui le informazioni anamnestiche non strutturate sono disposte in forma narrativa. La digitalizzazione della documentazione di riferimento ha agevolato i processi di riconoscimento della scrittura manuale e della loro successiva trascrizione attraverso una fase di addestramento sul software Transkribus. Mediante la selezione di un gruppo di documenti digitalizzati, in particolare delle sezioni di anamnesi, è stato possibile avviare una fase di addestramento per lo sviluppo di un modello capace di automatizzare il processo di trascrizione dei vari tipi di

redazione manuale delle cartelle. L'attività descritta ha delineato una metodologia che partendo dalla digitalizzazione e trascrizione di documenti clinici ha posto le basi per l'implementazione di tecniche di categorizzazione partendo dall'uso di categorie predefinite per la fase di training. Ciò ha permesso di associare a ogni frase ritenuta rilevante da parte degli esperti di dominio nelle sezioni anamnestiche della documentazione analizzata un tag semantico da mappare con le macro-aree del database interno della struttura. Un inserimento automatico delle informazioni anamnestiche di ogni paziente a partire dalle annotazioni semantiche sulle trascrizioni della documentazione clinica permetterebbe di ottenere un livello di precisione e accuratezza maggiore nell'inserimento delle informazioni e nel loro recupero.

## Riferimenti bibliografici

- Abbas, Asim et al. 2021. "Clinical Concept Extraction with Lexical Semantics to Support Automatic Annotation." *International Journal of Environmental Research and Public Health* 18 (20): 10564. <https://doi.org/10.3390/ijerph182010564>.
- Attardi, Giuseppe, Vittoria Cozza, and Daniele Sartiano. 2015. "Annotation and Extraction of Relations from Italian Medical Records." In *Proceedings of 6th Italian Information Retrieval Workshop*, Cagliari: CEUR Workshop Proceedings.
- Bozzi, Andrea. 2000. "Computer-aided recovery and analysis of damaged text documents." *CLUEB Manuali e antologie*.
- Hassanzadeh, Hamed, Anthony Nguyen, and Bevan Koopman. 2016. "Evaluation of medical concept annotation systems on clinical records." In *Proceedings of the Australasian Language Technology Association Workshop*, 15-24.
- How, Bong Chin and Kulathuramaiyer. 2004. "An empirical study of feature selection for text categorization based on term weightage." In *IEEE/WIC/ACM International Conference on Web Intelligence (WI'04)*, 599-602.
- Mykowiecka, Agnieszka, Małgorzata Marciniak, and Anna Kupść. 2009. "Rule-Based Information Extraction from Patients' Clinical Data". *Journal of Biomedical Informatics* 42 (5): 923-36. <https://doi.org/10.1016/j.jbi.2009.07.007>.

- Nockels, Joe, Paul Gooding, Sarah Ames, and Melissa Terras. 2022. "Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research". *Arch Sci* 22: 367–392. <https://doi.org/10.1007/s10502-022-09397-0>.
- Pasceri, Erika, Mérième Bouhandi, Claudia Lanza, Anna Perri, Valentina Laganà, Raffaele Maletta, Raffaele Di Lorenzo, and Amalia C. Bruni. 2023. "Neurodegenerative Clinical Records Analyzer: Detection of Recurrent Patterns Within Clinical Records towards the Identification of Typical Signs of Neurodegenerative Disease History". *JLIS.It* 14 (2):20-38. <https://doi.org/10.36253/jlis.it-522>.
- Pawar, Pratiksha Y and Shravan, H. Gawande. 2012. "A comparative study on different types of approaches to text categorization". *International Journal of Machine Learning and Computing*, 2(4): 423.
- Pinal, Pinal, Disha Davey, Vishal Panchal, and Parth Pathak. 2018. "Annotation of a large clinical entity corpus." In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2033-2042.
- Valacchi, Federico. 2006 *La memoria integrata nell'era digitale. Continuità archivistica e innovazione tecnologica*. Corazzano: Titivillus Edizioni.