

CNR-IRCrES Working Paper

Thesaurus e glossario ragionato della lingua della IA e della robotica



3/2024

Grazia Biorci

Direttore Giuseppe Giulio Calabrese

Direzione CNR-IRCrES
Istituto di Ricerca sulla Crescita Economica Sostenibile
Strada delle Cacce 73, 10135 Torino, Italy
Tel. +39 011 3977612
segreteria@ircres.cnr.it
www.ircres.cnr.it

Sede di Roma Via dei Taurini 19, 00185 Roma, Italy
Tel. +39 06 49937809 / Fax +39 06 49937808

Sede di Milano Via Corti 12, 20121 Milano, Italy
Tel. +39 02 23699501 / Fax +39 02 23699530

Sede di Genova Corso Ferdinando Maria Perrone 24, 16152 Genova, Italy
Tel. +39 010 6598798

Comitato Redazione

Giuseppe Giulio Calabrese, Grazia Biorci, Igor Benati, Antonella Emina, Serena Fabrizio, Lucio Morettini, Susanna Paleari, Anna Perin, Emanuela Reale, Secondo Rolfo, Andrea Orazio Spinello, Isabella Maria Zoppi.

✉ redazione@ircres.cnr.it
🌐 <https://www.ircres.cnr.it/produzione-scientifica/>

The Working Papers published by CNR-IRCrES represent the views of the respective author(s) and not of the Institute as a whole.

CNR-IRCrES Working Paper 3/2024



maggio 2024 by CNR-IRCrES

Thesaurus e glossario ragionato della lingua della IA e della robotica*

Thesaurus and reasoned glossary of AI and robotic language

GRAZIA BIORCI

CNR-IRCrES, Consiglio Nazionale delle Ricerche, Istituto di Ricerca sulla Crescita Economica Sostenibile, corso Ferdinando Maria Perrone 24, 16152 Genova, Italia

corresponding author: grazia.biorci@ircres.cnr.it

ABSTRACT

In the framework of the RAISE project, Spoke 5 - P6 Education, the online community STACY (Secondary TeAcher CommuniTY) was created in 2023, comprising teachers and researchers with the aim of fostering a productive exchange between the realms of research and education concerning the understanding and advancement of Artificial Intelligence (AI), Robotics, and Anticipatory Thinking. Within this framework, a study has been initiated to undertake metalinguistic analysis of the terminology used in the communication and portrayal of AI and Robotics in the Italian mainstream media. The aim is to enhance and strengthen linguistic resources conducive to effective shared communication within the specialized language of AI and Robotics. To this end, a metalinguistic reflection activity is proposed on a series of terms commonly utilized in the standard Italian language, related to the semantic field of AI and Robotics. This selection will be based on frequency analysis derived from a review of seven widely circulated Italian newspapers (*La Repubblica*, *Corriere della sera*, *Il Secolo XIX*, *Il Foglio*, *Il Fatto quotidiano*, *Corriere del Mezzogiorno*, *Liberio*). The findings will be presented in the form of a thesaurus, including a glossary of sector-specific terms, delineated across two distinct registers: the common language prevalent in generalist newspapers and magazines, and the specialized.

KEYWORDS: AI, sectorial language, Thesaurus.

DOI: attribuito dalla redazione

HOW TO CITE

Biorci, G. (2024). *Thesaurus e glossario ragionato della lingua della IA e della robotica* (CNR-IRCrES Working Paper 03/2024). Istituto di Ricerca sulla Crescita Economica Sostenibile. <http://dx.doi.org/10.23760/2421-7158.2024.003>

*RAISE per la Scuola. Spoke 5 - P6 Formazione. Capofila ITD CNR, Stefania Bocconi, Guglielmo Trentin.

INDICE

1.	INTRODUZIONE	3
2.	PERCHÉ UN THESAURUS?.....	4
3.	LA COSTRUZIONE DEL THESAURUS.....	5
4.	METODO DI SPOGLIO E DI ANALISI DEI DATI.....	6
5.	IMPLEMENTARE IL THESAURUS ATTRAVERSO GLI OUTPUT DI LABORATORI NELLE SCUOLE.....	10
6.	COME SI CONSULTA E COME SI USA QUESTO THESAURUS?.....	10
7.	RIFERIMENTI BIBLIOGRAFICI.....	15

1. INTRODUZIONE

In ambiente tecnico-scientifico e accademico si parla di Intelligenza Artificiale e di robotica da diversi decenni, ma finora tali tecnologie non avevano avuto un impatto reale sulla vita del pubblico generico. Concetti come IA e Robotica occupavano uno spazio all'interno della fantascienza o in avveniristici laboratori del mondo degli scienziati. Lo scenario per le cosiddette persone comuni è cambiato repentinamente nel marzo 2023, quando ChatGPT ha fatto irruzione nelle vite di tutti. Per la prima volta tante persone sono entrate in contatto diretto, e consapevole, con l'Intelligenza Artificiale e ne hanno intuito tutte le potenzialità, l'impatto e l'immensità.

C'è stato un momento di disorientamento generalizzato. La maggior parte delle persone era impreparata a questa rivoluzione culturale. Nessuno fino a quel momento aveva mai dovuto riflettere su termini come *machine learning*, reti neurali, *deep learning*, né aveva mai realizzato che molte tecnologie *soft*, entrate in punta di piedi nel quotidiano, come il riconoscimento facciale o i comandi vocali del cellulare, si basavano già su principi analoghi all'IA.

In quella primavera, tutti siamo stati travolti da un'ondata di terminologia specifica e nuova, il cui significato non sempre risultava chiaro. Sebbene intuitivamente si potesse indovinarne senso e funzione, alcune espressioni rimanevano piuttosto opache e il loro significato un po' vago. Questo ha stimolato la necessità e l'urgenza di acquisire in tempi brevi il nuovo lessico, che è diventato indifferibile conoscere e utilizzare.

Poiché si tratta di un linguaggio nuovo, che sottende concetti, tecnologia e oggetti nuovi – che rapidamente diverranno la quotidianità per l'intera popolazione nei prossimi decenni –, si ritiene possa essere utile iniziare a riflettere proprio sulle parole che concorrono alla narrazione/definizione dell'IA. Uno degli obiettivi di questo lavoro è mitigare possibili incomprensioni o significazioni erronee o parziali che spesso contribuiscono a un intendimento distorto o non funzionale della IA e della robotica. L'opacità di qualunque tipo di terminologia settoriale comporta, in molti casi, una cascata di ulteriori incomprensioni che possono generare false notizie, pregiudizi o diffidenze, aumentando in questo modo le ambivalenze e le importanti difficoltà di comunicazione che ogni novità comporta.

Con la proposta del progetto del Thesaurus di seguito esposta si vorrebbe, almeno dal punto di vista linguistico, offrire un panorama privo di ambiguità e fondato scientificamente in modo da permettere, in primo luogo ai formatori, ma anche a tutta la popolazione e a tutte le generazioni, la condivisione consapevole di un linguaggio comune relativo a un settore ancora poco conosciuto. Con queste premesse, si è pensato di iniziare proprio dalla lingua del giornalismo, la lingua che sa trasferire dalla fonte al lettore in modo comprensibile, senza cadere nelle tentazioni dei tecnicismi. Nel caso della robotica e della IA, il ruolo dei divulgatori scientifici e dei giornalisti è fondamentale: tali figure si pongono nei confronti dei lettori e del pubblico proprio come mediatori culturali di un mondo nuovo. In particolare, per questo tipo di tecnologie che impattano (e impatteranno) fortemente nella vita di ciascuno, la mediazione linguistica e culturale è fondamentale. Si possono intravedere, già al primissimo approccio, che esistono due livelli di linguaggio della comunicazione: la terminologia scientifica, specifica e tecnica, nota agli esperti, e il linguaggio comune, quotidiano e popolare dei non esperti del settore, pur comprendendo registri alti e dotti.

La lingua settoriale scientifica è sostanzialmente facile da reperire: è la lingua utilizzata nei convegni internazionali, nelle pubblicazioni scientifiche e nei report e negli articoli tecnici. Solitamente è una lingua condivisa a livello internazionale. Tuttavia, non è raro che esistano delle varietà anche all'interno della lingua settoriale, dettate principalmente dai diversi approcci a un dato fenomeno, dalle numerose scuole di pensiero, dalle università o dai gruppi di ricerca nei laboratori sparsi in tutto il mondo. La comunicazione può presentare delle criticità anche nelle scienze cosiddette dure.

Nel registro comune, nella lingua non settoriale, in che modo, in che linguaggio si può scrivere e parlare di IA? Certamente il serbatoio cui si attinge è quello della terminologia tecnico-

scientifico, spesso presa integralmente nella sua forma inglese. Talvolta, soprattutto nel tentativo di dare una definizione o di rendere comprensibile un concetto o un fenomeno nuovo, si utilizzano termini (vecchi e noti) di registro quotidiano, seppure alto, con una valenza intermedia tra l'eloquio della strada e quello scientifico: un linguaggio, quindi, pseudo-scientifico intelligibile e trasferibile senza troppe difficoltà a tutti i lettori e i parlanti. Questo tipo di linguaggio intermedio potrebbe rappresentare una proposta didattica nelle scuole primarie e secondarie di primo grado. Come nell'apprendimento di una lingua straniera, la lingua tecnologica della IA va in primo luogo acquisita dagli insegnanti, non ancora tutti nativi digitali o parlanti nella lingua tecnologica, i quali potranno a loro volta proporla in classe come materia di studio: un momento didattico per riflettere con gli allievi sui nuovi concetti di intelligenza, di evoluzione, di robotica, di etica e di socialità. Questo potrebbe contribuire ad appianare le incomprensioni intergenerazionali sulla conoscenza e l'uso delle tecnologie e potrebbe contribuire a mitigare l'opacità di alcune espressioni date per scontate, delle quali si presume tutti debbano già sapere il significato.

2. PERCHÉ UN THESAURUS?¹

Con questi obiettivi in mente, si è iniziato a progettare la raccolta e la sistematizzazione della lingua settoriale dell'IA e della Robotica da rendere fruibile come strumento didattico. Si poteva immaginare un semplice glossario con le definizioni scientifiche ragionate o rese più comprensibili agli utenti attraverso un processo di riscrittura, oppure si poteva immaginare di affiancare al glossario ragionato un dizionario analogico, un Thesaurus². Quest'ultimo progetto è sembrato coerente con gli obiettivi del progetto RAISE, in quanto risponde sia alle esigenze di formazione per insegnanti e formatori sia alla messa in atto di buone pratiche di inclusione intergenerazionale su argomenti relativi alle tecnologie avanzate.

Il progetto ha preso dunque avvio. La prima decisione da prendere è stata quella di definire le fonti dalle quali acquisire i testi da sottoporre ad analisi linguistica. Poiché l'obiettivo è quello di costruire uno strumento didattico, si è optato per una ricerca di testi che abbiano un grande impatto comunicativo, garantendo però autorevolezza e serietà. La scelta è ricaduta sui quotidiani a diffusione nazionale.

L'idea era di iniziare a raccogliere le diverse definizioni e citazioni dell'Intelligenza Artificiale e dei suoi intorni, linguistici e concettuali, nella lingua della stampa quotidiana che rappresenta ancora, almeno per molte persone adulte, il mezzo privilegiato per conoscere ciò che avviene nel mondo e nella società. Inoltre, il linguaggio della stampa, per definizione, raggiunge quella vasta parte della popolazione che usa termini non settoriali o limitatamente settoriali e, eventualmente, ne fornisce delle spiegazioni in maniera comprensibile alla maggioranza. In ogni caso, la stampa gode ancora di una certa autorevolezza e i giornalisti sono professionalmente tenuti alla verifica e all'attendibilità delle fonti. Non solo, la lingua della stampa raggiunge tutte le persone e, a differenza della letteratura scientifica, può essere un misuratore della consapevolezza sia dei giornalisti sia dei lettori, in particolare per tutto ciò che riguarda le nuove tecnologie e gli scenari futuri che queste possono prefigurare, con tutte le emozioni e le ambivalenze che ne derivano.

Si è deciso così di creare un *corpus linguistico* composto dalla lingua prodotta dai giornalisti per riportare notizie e descrivere l'IA e la robotica nelle pagine dei quotidiani, adottando la metodologia della ricerca *corpus-driven* (Sinclair, 1991; Tognini-Bonelli, 2001).

Dal gennaio 2023 sono stati raccolti manualmente gli articoli pubblicati sui giornali che contenevano gli argomenti dell'Intelligenza Artificiale e della robotica, anche se non esplicitato nel titolo: una sorta di rassegna stampa su sette quotidiani a tiratura nazionale. La scelta dei quotidiani ha comportato la gestione di alcuni problemi metodologici imprescindibili. Quali

¹ <https://stacy-community.it/index.php/iniziativa/thesaurus-e-glossario-ragionato-della-lingua-della-ai-r-2>

² Un dizionario analogico, o un Thesaurus è un dizionario in cui le parole non sono definite per il loro significato, ma si trovano "in compagnia" delle altre parole con le quale formano campi semantici specifici. Per esempio, la parola "chiodo" si trova nella voce "martello".

giornali? Quale distribuzione sul territorio nazionale? Come reperire file in formato testo per poter sottoporre il corpus ad analisi testuale e linguistica?

Senza pretendere di essere ineccepibili dal punto di vista formale e metodologico, ci si è orientati per un processo fattibile di raccolta e indagine, per poter iniziare a mettere insieme da subito i dati testuali. Sulla piattaforma MLOL³ della città metropolitana di Genova si sono potute verificare le testate disponibili alla lettura e al download di articoli in formato txt o word; si è potuta verificare la distribuzione territoriale dei quotidiani stessi e si sono definiti i criteri di scelta.

- La maggior parte dei quotidiani – cinque su sette – ha carattere nazionale: *La Repubblica*, *Corriere della sera*, *Il Foglio*, *Il Fatto quotidiano*, *Libero*.
- Due hanno carattere più locale, ma con respiro territoriale ampio: *Il Secolo XIX* (che copre il territorio della Liguria e del basso Piemonte) e *Il Corriere del Mezzogiorno* (Puglia, Campania e Sicilia).

La scelta è stata dettata soprattutto dal fatto che gli articoli contenuti in tali testate possono essere scaricati gratuitamente senza autorizzazioni speciali. Va da sé che, trattandosi di materiale soggetto a copyright, l'utilizzo che ne è stato fatto prevedeva l'uso del materiale solo per la ricerca e l'analisi testuale.

Durante il 2023 sono stati così collezionati, divisi per testata e per data, 154 articoli contenenti argomenti relativi all'Intelligenza Artificiale e alla robotica, che costituiscono un piccolo *corpus* di 165.252 forme (*tokens*), ovvero 18.466 parole diverse (*types*).

3. LA COSTRUZIONE DEL THESAURUS

Con questo materiale si è avviata l'analisi testuale finalizzata alla costruzione del Thesaurus. Un Thesaurus è concepito come una sorta di dizionario concettuale nel quale le parole sono collocate secondo un ordine logico, collegate a campi semantici o a serie e catene nomenclatorie. È uno strumento linguistico molto pratico, meno artificiale e astratto rispetto a un ordine alfabetico. La sua peculiarità risiede nell'espressione delle componenti culturali, antropologiche e tradizionali di ciascun termine. Infatti, si registrano le parole in associazione ai loro legami semantici, alle loro collocazioni abituali nella lingua dell'uso, alle co-occorrenze più frequenti e alla loro importanza nel testo. Interessava preparare un simile oggetto linguistico per capire e raccogliere tutte le percezioni, le idee, le paure e i mille scenari suscitati e risvegliati dal contatto con l'Intelligenza Artificiale, che, nell'ultimo anno, ha prepotentemente scosso e messo in discussione la vita del pubblico generico/cittadini comuni/non addetti ai lavori.

Proprio per promuovere una consultazione agevole del nostro Thesaurus, si è ritenuto utile realizzare contestualmente un piccolo glossario, nel quale vengono fornite ai termini specifici di alto uso, come *ChatGPT*, *deep learning* e *machine learning*, anche le definizioni scientifiche maggiormente condivise.

Per questa sua particolarità, il Thesaurus che è in corso di realizzazione è *sui generis*: si tratta infatti di una raccolta organizzata di parole, complete di definizione scientifica, che gravitano intorno ai concetti di IA e di robotica. Ad ogni voce del Thesaurus, quindi, corrisponde sia una spiegazione tecnico-scientifica sia una nuvola di altri termini che costituiscono l'intorno linguistico del termine esaminato. Tali parole sono associate al termine in esame attraverso collegamenti di varia natura: carattere emotivo, carattere tecnico, uso frequente, collocazione, perché forme idiomatiche o per solidarietà sintagmatica. Il risultato è uno strumento in cui le voci sono spiegate su due registri linguistici: il registro della lingua comune e divulgativa (nei quotidiani, nelle riviste generaliste e nei contesti d'uso quotidiano); il registro della lingua specialistica tecnico-scientifica nei suoi contesti d'uso specifici.

Le voci in entrata sono i termini che nel *corpus* risultano avere la frequenza maggiore. Queste, lemmatizzate, costituiscono un elenco di forme concettualmente legate alla IA e alla robotica.

³ <https://bigmet.medialibrary.it/home/index.aspx>

Alcune appartengono alla nomenclatura specifica della IA (*ChatGPT*, *deep learning* e *machine learning*), altre invece sono costituite da termine più generici, seppure molto frequenti: *macchina*, *sviluppo*, *rischio*, *lavoro*. Queste ultime sono le *altre* parole associate concettualmente alla voce in entrata: si distinguono in co-occorrenze, collocazioni, associazioni ricorrenti e non solo si presentano con una certa frequenza, ma assumono all'interno dell'intero *corpus* un valore significativo quanto a salienza e rilevanza.

Fra le parole legate concettualmente alla voce in entrata, molte presentano connotazioni emotive.

4. METODO DI SPOGLIO E DI ANALISI DEI DATI

La fase di raccolta dei dati è durata un anno e ha comportato la consultazione giornaliera sul portale MLOL⁴ dei sette quotidiani scelti, effettuando la rassegna stampa orientata a IA e robotica. Questo compito è stato eseguito a mano, anche perché si è privilegiato lo sguardo critico di chi ricercava contenuti rispetto alla presenza o meno delle parole chiave – IA, AI e robot(ica) – nei titoli dei diversi articoli. Va notato che sono stati considerati gli articoli apparsi in qualunque pagina dei quotidiani: dalle pagine della scienza, da quelle economiche o di cronaca e, spesso, dagli articoli di fondo o da quelli di cultura e società della terza pagina.

In 11 mesi, sono stati raccolti 154 articoli: una media di 3 alla settimana, che dà una media di 1 articolo ogni 2 quotidiani.

È stato interessante osservare la distribuzione degli articoli nei diversi quotidiani. Fra i più attenti alle nuove proposte tecnologiche della scienza, ma con affondi importanti anche sul piano economico, ci sono *La Repubblica* e il *Corriere della Sera*, che sono usciti anche con inserti dedicati, in particolare in occasione dell'Italian Tech Week che si è tenuta a Torino nel settembre 2023. Ancora piuttosto presente e con diverse rubriche dedicate anche *Il Secolo XIX*, certamente platea privilegiata per l'Istituto Italiano di Tecnologia (IIT) che ha sede, appunto, sul territorio genovese. *Il Foglio*, *Il Fatto Quotidiano*, *Liberò* e il *Corriere del Mezzogiorno*, benché siano stati meno assidui nella comunicazione dei progressi delle nuove tecnologie, hanno concentrato l'interesse sulle relative problematiche e, in particolare, sulle ricadute sul piano politico-sociale, sul lavoro e sulla vita delle persone.

Gli articoli, divisi per testata e data, sono stati processati per lo spoglio automatico e la rilevazione dei dati statistici dall'applicativo open-source AntConc®⁵.

Si sono così ottenute le liste delle parole più frequenti in tutto il *corpus*.

Da questo gruppo sono state estratte solo quelle occorrenze che hanno un valore semantico, ossia verbi, sostantivi e aggettivi. Non si è ritenuto necessario effettuare una lemmatizzazione del *corpus*, poiché il numero di parole diverse non raggiungeva le ventimila unità, a fronte di una raccolta di 165000 forme. Da un punto di vista statistico, tali numeri affermano una varietà lessicale abbastanza bassa. Per il nostro *corpus*, la formula $T=V/N$ produce il valore $18466/165252 = 0,11$. Questo coefficiente denota una certa ricorsività e una bassa varietà lessicale, ma questo è un fenomeno tipico dei testi scientifici⁶.

Dal *corpus* è stata estratta la lista di parole diverse con relative frequenze assolute (18466). Sono state successivamente isolate tutte le parole diverse con frequenza maggiore o uguale a 3 ottenendo un elenco di 6116 occorrenze⁷. Da questo elenco sono state scartate le forme

⁴ <https://www.medialibrary.it/home/index.aspx>

⁵ <https://www.laurenceanthony.net/software/antconc/>

⁶ Tuttavia, è solo analizzando qualitativamente e quantitativamente i termini che circondano le voci più frequenti negli articoli su IA che si possono trarre conclusioni. Infatti, accanto alla tendenza ad affrontare i temi legati alla IA e alla robotica utilizzando un lessico di carattere più scientifico e asettico, si rilevano cotesti di Intelligenza Artificiale e robotica in cui ricorrono in modo ripetitivo descrittori emotivi quali *distruzione*, *annientamento*, *pericolo*, *sopravvivenza*.

⁷ Nell'analisi delle co-occorrenze sono state considerate anche le parole con frequenza inferiore a 3, principalmente quando queste erano associate a termini molto frequenti. Nella colonna collocazioni, i lemmi senza numero sono quelli con frequenza uguale a 1.

grammaticali (articoli, preposizioni, congiunzioni, pronomi personali, aggettivi possessivi, aggettivi indefiniti, aggettivi e pronomi dimostrativi e le forme coniugate dei verbi essere e avere). Sono state raccolte le forme con valore semantico con frequenza superiore o uguale a 3, per un valore complessivo di circa 500 termini diversi che, lemmatizzati manualmente, hanno prodotto una lista di circa 160 lemmi. Grazie all'applicativo AntConc®, per ciascuno di questi si sono estratti i contesti, i cluster più ricorrenti – composti da 2 a 5 forme –, le collocazioni – comprendenti almeno 3 forme (Carrol & Conklin, 2019). Si sono potuti così ricavare e contare i termini che, nel nostro *corpus*, sono associati alle occorrenze più numerose o le accompagnano.

L'*output* di tale analisi è un file excel composto da quattro colonne.

Nella prima colonna è riportata la definizione tecnica o scientifica o pseudoscientifica del lemma o del sintagma della seconda colonna, sul quale sono state effettuate le analisi testuali. Nella terza colonna è espressa la frequenza del lemma o del sintagma all'interno del corpus. Nella quarta colonna sono elencati in ordine decrescente i sostantivi, gli aggettivi e i verbi nella loro forma lemmatizzata, ossia i sostantivi al singolare, gli aggettivi al maschile e i verbi all'infinito, con il valore della loro attestazione all'interno del corpus in relazione alla parola della seconda colonna.

La definizione scientifica presente in ogni casella della prima colonna termina con l'indicazione della fonte dalla quale è stata ottenuta. La maggior parte delle definizioni proviene da pubblicazioni scientifiche del settore tradotte in italiano; altre provengono dalla sezione Neologismi di autorevoli enciclopedie online; altre ancora da dizionari online e cartacei; altre, infine, per prova o provocazione, da ChatGPT.

Nella fase iniziale del progetto, sono stati estratti le parole o i sintagmi molto frequenti negli articoli relativi a IA e robotica che in qualche modo necessitano di essere spiegati e descritti al pubblico dei non esperti. Infatti, la redazione di questo Thesaurus è stata pensata per sciogliere quelle ambiguità di senso che spesso rimangono per chi affronta le prime letture.

In particolare, si riteneva fosse importante immaginare uno strumento didattico utile per i formatori, incrementabile grazie alla partecipazione generale e all'ulteriore ricerca di nuove parole su IA e di nuove suggestioni e indagini all'interno del lessico della stampa quotidiana.

Dopo qualche mese dall'ultima redazione del Thesaurus, sono emersi altri termini che andrebbero aggiunti all'interno delle voci finora selezionate. Nella prima lettura del *corpus*, non si erano prese in considerazione alcune parole relative alla IA connotate in modo fortemente negativo, come *preoccupazione*, *pericolo*, *medicina*, e altre che, a distanza di tempo, pare non possano essere ignorate. Da una parte, la loro presenza all'interno del corpus indica la loro salienza. *Pericolo* e il suo aggettivo *pericoloso/a* hanno frequenza assoluta complessiva pari a 85 nell'intero *corpus*. *Minacce* e *morte* compaiono ciascuna 11 volte e, osservando i contesti in cui sono immerse, appaiono fra loro vicine o contigue a espressioni come *dell'umanità*, *del genere umano*, *della civiltà*.

Di seguito, si propongono alcuni grafici esemplificativi nei quali sono individuabili le co-occorrenze riscontrate con maggiore frequenza e salienza nell'intorno del termine con alta frequenza nel *corpus*.

Figura 1. Intelligenza Artificiale

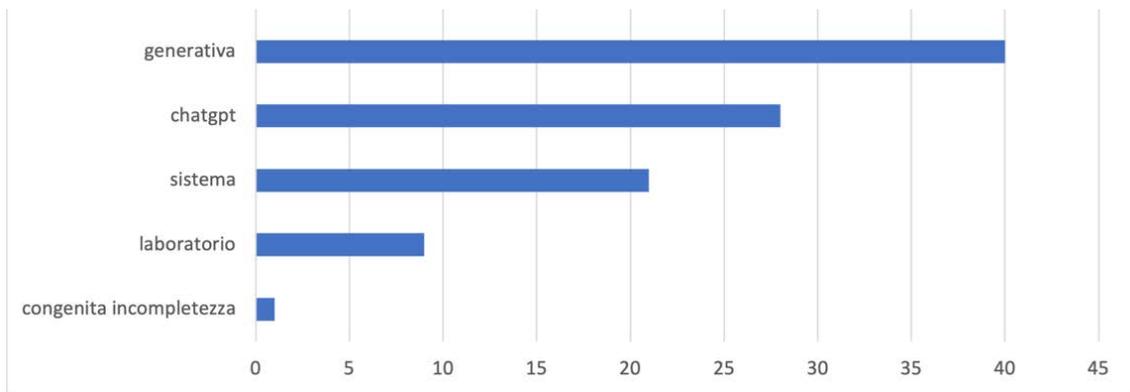


Figura 2. Algoritmo

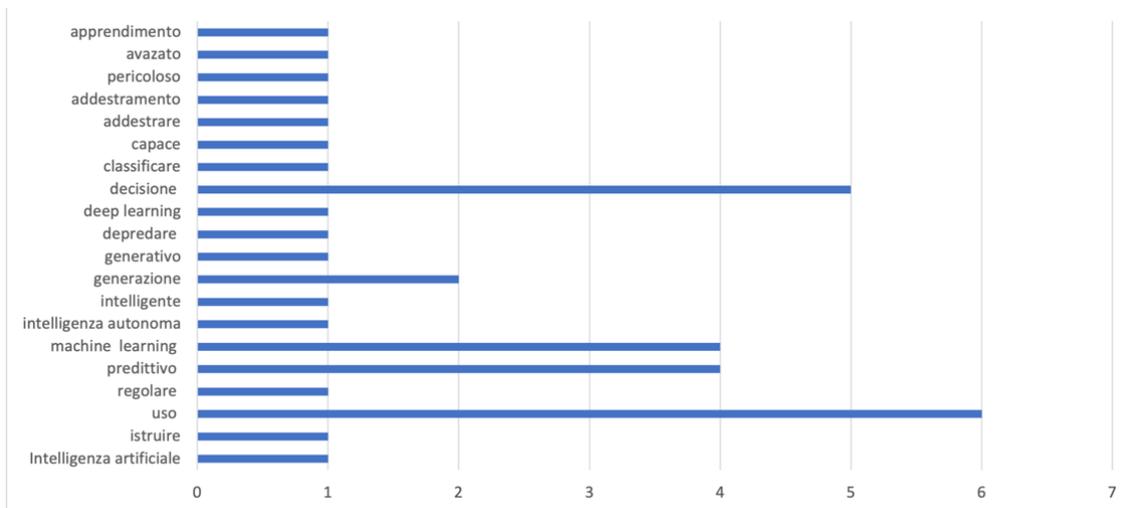


Figura 3. Machine learning

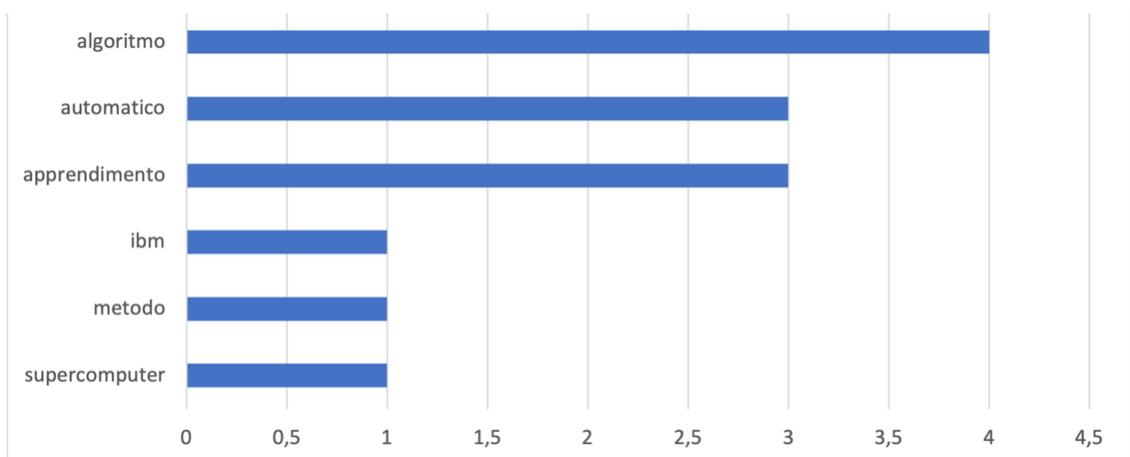


Figura 4. AI

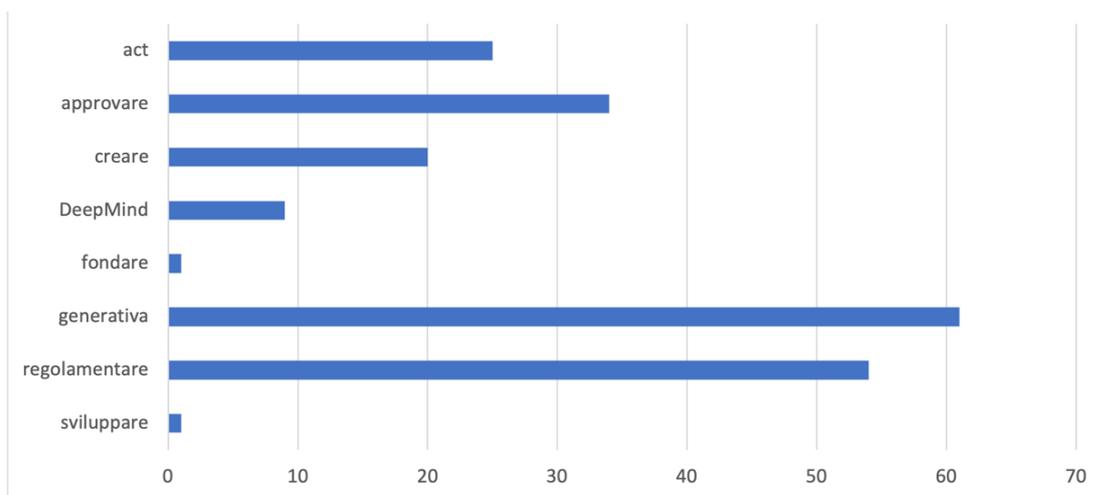


Figura 5. Addestrare

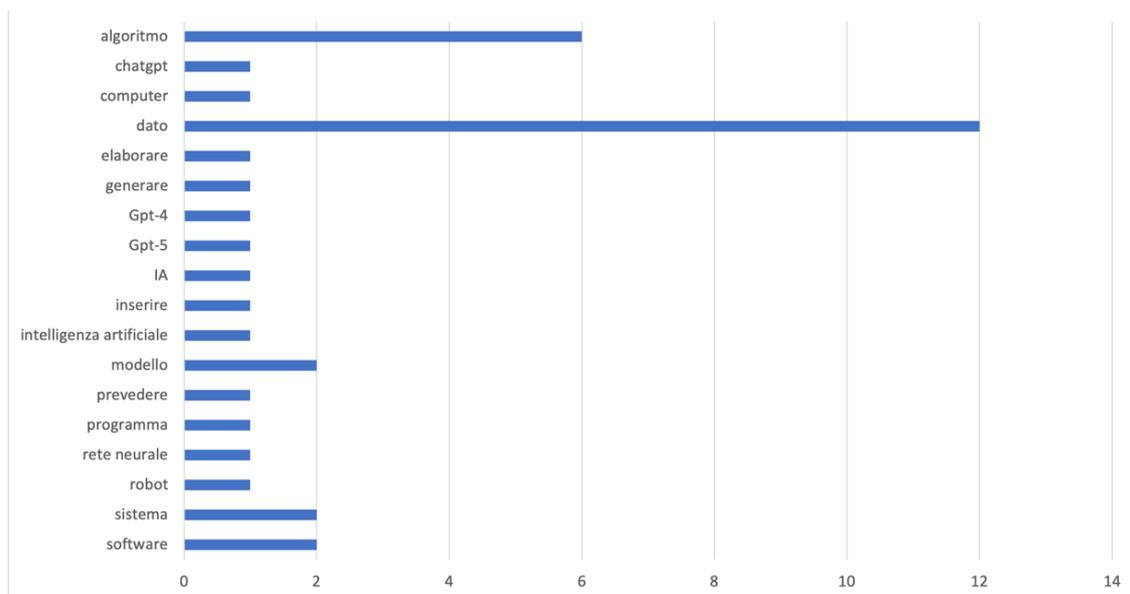
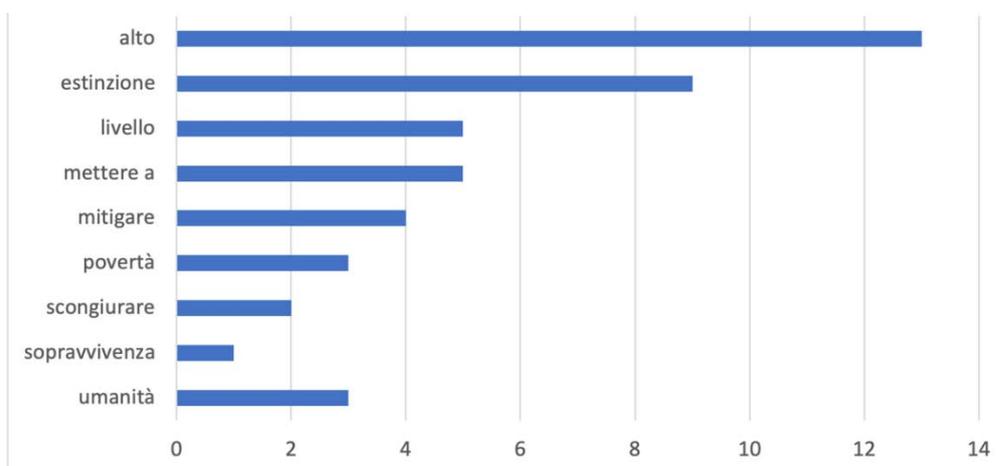


Figura 6. Rischio



5. IMPLEMENTARE IL THESAURUS ATTRAVERSO GLI OUTPUT DI LABORATORI NELLE SCUOLE

Il *corpus* potrebbe essere ancora studiato e analizzato sotto molti aspetti: dal conteggio delle co-occorrenze alla *sentiment analysis*, fino alla costruzione di campi semantici specifici. L'obiettivo principale del lavoro è ampliare e consolidare strumenti linguistici adeguati a un'efficace comunicazione condivisa della lingua settoriale della IA e della robotica attraverso la costruzione di uno strumento da utilizzare all'interno di laboratori didattici, per una interoperabilità linguistica e concettuale fra insegnanti e allievi.

Si è pensato di introdurre il Thesaurus all'interno di laboratori previsti sia per formatori sia per studenti, proponendo attività di riflessione metalinguistica su una serie di termini comunemente utilizzati nella lingua standard italiana e presenti nell'ambito semantico della IA e della robotica.

L'attività prevede un lavoro di gruppo: la fase iniziale di brainstorming e di produzione libera di terminologia nota ai partecipanti (lingua d'uso) è seguita da una rappresentazione concettuale delle relazioni linguistiche intorno alle parole/concetti IA o robotica. In seguito, si confrontano i risultati ottenuti dal *brainstorming* con i risultati tratti dallo spoglio della rassegna stampa (lingua comune divulgativa). Per ogni nuovo termine il gruppo costruisce una definizione partecipata secondo le conoscenze, da confrontare e valutare in relazione con gli altri termini e in particolare con la nomenclatura scientifica o con accezione tecnico-scientifica (lingua specialistica).

In questo modo, il Thesaurus cresce nel numero, nella varietà lessicale e nella varietà dei termini settoriali su diversi livelli: 1) nella lingua comune e divulgativa dei quotidiani e delle riviste generaliste; 2) nella lingua specialistica scientifica e 3) nella lingua d'uso dei parlanti.

6. COME SI CONSULTA E COME SI USA QUESTO THESAURUS?

È ancora allo studio l'interfaccia per l'utilizzo del Thesaurus⁸. In linea teorica l'utente di un thesaurus è interessato a conoscere non tanto il significato di una parola quanto il contesto e le solidarietà sintagmatiche nei quali la parola è immersa.

In questo Thesaurus, il contesto nel quale il termine è immerso e connesso è quello derivante principalmente dallo spoglio dei testi della rassegna stampa che, una volta realizzati i laboratori, sarà ampliato grazie ai contributi delle persone che vi avranno partecipato.

A titolo di esempio, si trascrivono di seguito alcune righe del file excel nella versione di lavoro finora realizzata. Si potranno leggere nella colonna A le definizioni, tratte da fonti scientifiche, delle voci in entrata segnalate nella colonna B; nella colonna C è riportata la loro frequenza assoluta nel *corpus*; nella colonna D, si leggono in ordine decrescente e con la loro frequenza nel *corpus* le parole che costituiscono le collocazioni, i sintagmi ricorrenti e le co-occorrenze salienti relative alla voce in entrata.

Tabella 1. Trascrizione di alcune righe del file excel nella versione di lavoro per la redazione del Thesaurus

A definizione scientifica	B termine o locuzione	C frequenza	D collocations
	addestrare	72	algoritmo 6, dato 12, rete neurale, modello 2, computer, programma, sistema 2, software 2, robot, IA, Intelligenza

⁸ Ringrazio le colleghe e i colleghi di ITD Stefania Bocconi, Francesca Talamini, Guglielmo Trentin e Giovanni Caruso per essersi resi disponibili a trovare (e realizzare) un'interfaccia di gestione del Thesaurus adatta alla fruizione in rete.

			Artificiale, ChatGPT, Gpt-5, Gpt-4; generare, elaborare, prevedere, inserire
Artificial Intelligence	AI	391	act 25, generativa 61, DeepMind 9; sviluppare, creare 20, approvare 34, regolamentare 54, fondare
1. TS st.mat. nel Medioevo, sistema di calcolo con cifre arabe 2a. TS mat. insieme di regole per la risoluzione di un calcolo numerico; gener., procedimento matematico TS inform. insieme di regole che forniscono una sequenza di operazioni atte a risolvere un particolare problema 2b. CO estens., metodo o procedimento sistematico per risolvere un problema https://dizionario.internazionale.it/parola/ algoritmo	algoritmo	164	Intelligenza Artificiale, avanzato, predittivo 4, capace, generativo, intelligente, intelligenza autonoma, decisione 5, apprendimento, deep learning, machine learning 4, pericoloso, addestramento, generazione 2, uso 6; istruire, classificare, depredare, addestrare, regolare
	allucinazione	9	risposta 3, controllata 2
Un'API, Application Programming Interface o interfaccia di programmazione delle applicazioni, è un insieme di regole definite che consente ad applicazioni diverse di comunicare l'una con l'altra. https://www.ibm.com/it-it/topics/api	API (interfaccia di programmazione di una applicazione)	2	interfaccia, programmazione
1. CO l'apprendere, l'imparare TS psic. processo di acquisizione di conoscenze, di informazioni o di comportamenti 2. TS etol. condizionamento del comportamento degli animali. https://dizionario.internazionale.it/parola/apprendimento	apprendimento	65	automatico 13, investigativo 4, profondo 2, autonomo 2, supervisionato 2, umano
Definita, nel tempo, come AI forte, le sue caratteristiche si sono differenziate dall'AI debole (chiamata anche AI stretta) non essendo quest'ultima concepita per eseguire capacità cognitive e personalità simili a quelle umane, bensì limitata all'uso del software per studiare o eseguire specifici compiti di problem solving o ragionamento pre-appresi (sistemi esperti). https://www.ai4business.it/intelligenza-artificiale/agi-intelligenza-artificiale-generale-cose-come-si-puo-utilizzare/	Artificial General Intelligence (AGI)	23	forma avanzatissima 2, ossessione, nuova frontiera 2, umanità 4; replicare 4
1. AU che è prodotto, ottenuto a imitazione della natura grazie a un procedimento tecnico https://dizionario.internazionale.it/parola/artificiale	artificiale	861	intelligenza 861, generativa 63, ChatGPT 12, generale 2, applicata 2, etica 2, software 6, sistema 10, istruzione 9, conversazionale, Google 2,

			sviluppata 2; provocare 2, aumentare 3, ottimizzare i processi 2, sostituire 2
	intelligenza	939	artificiale 834, umana 38, generativa 41, naturale 17, non 21, applicata 9, sistemi di 22, ChatGPT 11, sviluppo 10, organoide 2, sintetica 2, autonoma, biologica, priva (di coscienza, sensazioni basilari, dolore, piacere), frutto di 13, superiore, inferiore, disumana, senziente; sostituire 6, creare 2
L'Intelligenza Artificiale è una disciplina dell'informatica che si occupa di creare macchine in grado di imitare le capacità dell'intelligenza umana attraverso lo sviluppo di algoritmi che consentono di mostrare attività intelligente [...] Nella sua accezione puramente informatica, l'Intelligenza Artificiale potrebbe essere classificata come la disciplina che racchiude le teorie e le tecniche pratiche per lo sviluppo di algoritmi che consentano alle macchine (in particolare ai 'calcolatori') di mostrare attività intelligente, per lo meno in specifici domini e ambiti applicativi [...] quando si parla di capacità di ragionamento e apprendimento automatico sulla base dell'osservazione spesso si incappa nell'alveo del Cognitive Computing che va però inteso come l'insieme delle piattaforme tecnologiche basate sulle discipline scientifiche dell'Intelligenza Artificiale (tra cui Machine Learning e Deep Learning) e il Signal Processing (la capacità di elaborare i segnali). https://www.ai4business.it/intelligenza-artificiale/intelligenza-artificiale-cose/	Intelligenza Artificiale	766	ChatGPT 28, generativa 40, sistema 21, laboratorio 9, congenita incompletezza, non autosufficienza (in) grado 22, applicata 5; è (40), non è (22), offrire 4, fornire 2
I modelli linguistici di grandi dimensioni, tra cui ChatGPT, GPT-4 e altri, fanno esattamente una cosa: prendono in considerazione un gruppo di parole e cercano di indovinare quale parola dovrebbe venire dopo. Se questo è "ragionare" o "pensare", allora è solo una forma molto specializzata. https://medium.com	Large Language Models, LLM	20	LLM 2, modello 2, grande 2, linguistico 2, preesistente, minaccia, generale; rielaborare, memorizzare
Il termine "learning" in inglese si riferisce al processo di acquisizione di conoscenze, competenze, abilità o comportamenti attraverso l'esperienza, lo studio o l'addestramento. In ambito più specifico, il termine è spesso associato all'apprendimento nel contesto dell'istruzione o dell'adattamento a nuove situazioni. cit. ChatGPT 12 dic. 2023.	learning	48	algorithms, machine- 26, deep- 17, apprendimento 7, rete, reinforcement 3, metodo 3, automatico 3,

			profondo, rinforzo 2, sbrigativo, supercomputer; funzionare
	macchina	313	uomo 24, intelligente 28, macchina 16, romanzo 5, letteraria 5, senziante 4, pensante 3, rapporto 6; imparare 5, diventare 5, ricordare 4, svolgere, scrivere 4, rubare 3
<p>Il Machine learning (ML) insegna ai computer e ai robot a fare azioni e attività in modo naturale come gli esseri umani o gli animali: imparando dall'esperienza (o meglio, attraverso programmi di apprendimento automatico).</p> <p>In sostanza, gli algoritmi di Machine learning usano metodi matematico-computazionali per apprendere informazioni direttamente dai dati, senza modelli matematici ed equazioni predeterminate.</p> <p>Il Machine learning consente ai computer di abbinare algoritmi e tecnologie avanzate per prendere decisioni migliori, più rapidamente e più accuratamente. Uno dei principali vantaggi dell'apprendimento automatico è la capacità di modellare i dati e prendere decisioni in tempo reale.</p> <p>Ad esempio, un sistema di Machine learning può essere utilizzato per riconoscere le parole pronunciate da una persona o scansionare immagini che contengono oggetti specifici.</p> <p>Può anche essere utilizzato per prendere decisioni aziendali complesse, come la gestione delle risorse umane o l'analisi predittiva dei dati.</p> <p>Gli algoritmi di Machine learning sono una parte importante della tecnologia IA che gli imprenditori stanno adottando per aumentare la produttività, ridurre i costi e migliorare la loro capacità di prendere decisioni.</p> <p>Gli algoritmi Machine Learning possono essere usati per interpretare dati complessi e identificare tendenze, modelli e relazioni che possono essere utilizzati per prendere decisioni strategiche.</p> <p>Ad esempio, l'utilizzo di algoritmi machine learning nella sicurezza informatica aiuta a riconoscere le minacce come malware, intrusioni informatiche e frodi.</p> <p>https://www.ia4business.it</p>	Machine Learning	23	apprendimento 3, automatico 3, algoritmo 4, supercomputer, metodi, IBM
	manipolare	10	in grado 3, credulità, social 3, menti, fotografia; fuorviare, riprodurre
<p>I media sintetici sono qualsiasi tipo di video, immagini, oggetti virtuali, suoni o parole prodotti dall'Intelligenza Artificiale (o con l'aiuto di essa).</p> <p>https://www.digitalworlditalia.it/tecnologie-emergenti/intelligenza-artificiale/media-sintetici-trasformeranno-la-tecnologia-151555</p>	media sintetici	2	deepfake, richiesta, contenuto
	mente	41	umana 10, sano 3, estesa 3, banalità

			2, ragionamento 2, digitale
Midjourney, infatti, è una start-up che utilizza l'Intelligenza Artificiale per creare opere d'arte con risultati molto spesso sbalorditivi. Midjourney è un algoritmo di Intelligenza Artificiale text to image, ossia capace di generare splendide immagini in base a delle istruzioni di testo. https://www.aranzulla.it/	midjourney	53	Dall-E 9, Stable Diffusion 12, Stability AI, Deviantart, aggiornamento 3, immagini 6, De Chirico 2, mani; generare 3
	migliorare	59	rendimento 2, produttività 2, prodotti 3, performance 3
	minaccia	35	umanità 5, esistenziale 3, commerciale, disoccupazione, tecnologica; rappresentare 2
	minare	4	democrazie 2, società democratiche, stabilità, conflitti; sovvertire, aggravare
	mitigare	7	estinzione 4, rischio 6, dannoso
	modello	187	linguistico 31, apprendimento 10, Large Language Model 6, Transformer 3, statistico 4, educativo 4, Intelligenza Artificiale generativa 9, gpt 9, business 5; costruire 5, allenare 3.
	mondo	230	tutto il 25, fine del 10, intero 5, nuovo 13, reale; salvare 2, spiegare 2, rivoluzionario 2

I contesti presi in considerazione al momento sono questi, ma si potrebbe ampliare il *corpus* inserendo una scelta di testi derivanti da altri media, come trascrizioni di spezzoni di film, manga, canzoni o video che riguardano l'IA e la robotica. Potrebbe anche essere molto interessante analizzare la variazione diacronica della lingua settoriale, confrontando la terminologia del passato, tratta da letteratura scientifica, letteraria o divulgativa, con quella attuale, in un arco temporale di n anni. Ancora, potrebbero essere utili alla crescita del *corpus* le trascrizioni di interviste sia a cittadini di diverse età, condizione e scolarizzazione sia a scienziati e divulgatori scientifici. Ciò consentirebbe di osservare anche la variazione diastratica di questo particolare ambito semantico.

Il Thesaurus su IA e robotica, quindi, è un progetto incrementabile e in evoluzione, arricchito via via dal contributo di docenti e studenti e da vari eventi live.

Anche la parte di glossario tecnico, già consolidata e sostanzialmente condivisa fra i membri della comunità scientifica, è destinata a crescere e a modificarsi continuamente seguendo l'evoluzione e la specializzazione delle conoscenze e delle competenze.

7. RIFERIMENTI BIBLIOGRAFICI

Carrol, G., Conklin, K. (2020). Is All Formulaic Language Created Equal? Unpacking the Processing Advantage for Different Types of Formulaic Sequences. *Language and Speech*, 63(1), pp. 95-122.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at work*. John Benjamin Publishing Company.

CNR-IRCrES Working Papers

2024

N. 2/2024 [Comprendere le trasformazioni di San Lorenzo, Roma. Un tentativo di ordine concettuale tra gentrificazione e politica urbana](#). Antonia Astore, Luca Tricarico.

N. 1/2024 [How to measure the social acceptability of alternative environmental management solutions in wetlands and other ecosystems](#). Lisa Sella, Francesca Silvia Rota, Nicola Pollo.

2023

N. 5/2023 [Laboratori creativi sperimentali a distanza per la valutazione](#). Isabella Maria Zoppi, Antonella Emina.

N. 4/2023 [Eco-innovations and labor in the European automotive industry: an econometric study](#). Anna Novaresio.

N. 3/2023 [Climate change adaptation planning: tools and methods for effective and sustainable decisions](#). Filippo Frascini, Marco Pregolato.

N. 2/2023 [The effect of propensity to savings on rate of profit](#). Mario De Marchi.

N. 1/2023 [UNESCO o non UNESCO? Quando il brand incentiva il turismo: il caso dei paesaggi vitivinicoli delle Langhe e del Monferrato](#). Orsola Borsani, Greta Falavigna.

2022

N. 5/2022 [Logit statico, Logit dinamico e modelli hazard](#). Franco Varetto.

N. 4/2022 [Evolution of Deep Learning from Turing machine to Deep Learning next generation](#). Greta Falavigna.

N. 3/2022 [A Simulation Model of Technology Innovation of a Territory](#). Angelo Bonomi.

N. 2/2022 [Technology and Environmental Policies](#). Angelo Bonomi.

N. 1/2022 [Le donne marittime: fra stereotipi di genere, discriminazioni e scarse opportunità occupazionali](#). Barbara Bonciani, Silvia Peveri.

2021

N. 9/2021 [Management of open access research infrastructures in large EU projects: the "CultureLabs" case](#). Andrea Orazio Spinello, Danilo Giglito, Eleanor Lockley.

N. 8/2021 [Francia-Italia: parole in campo. Intorno alla narrazione del Campionato del Mondo di calcio donne](#). Antonella Emina.

N. 7/2021 [Covid-19 e rischio di insolvenza: il punto di vista del mercato azionario](#). Franco Varetto.

N. 6/2021 [Institutional efficiency and budget constraints: a Directional Distance Function approach to lead a key policy reform](#). Greta Falavigna, Roberto Ippoliti.

N. 5/2021 [Different waves and different policy interventions in 2020 Covid-19 in Italy: did they bring different results?](#). Mario Nosvelli.

N. 4/2021 [On Search of a General Model of Technology Innovation](#). Angelo Bonomi.

N. 3/2021 [Design and implementation of a web survey on the effects of evaluation on academic research](#). Andrea Orazio Spinello, Emanuela Reale, Antonio Zinilli.

N. 2/2021 [An online survey on the effects of agile working in Italian Public Research Organisations](#). Serena Fabrizio, Valentina Lamonica, Andrea Orazio Spinello.

N. 1/2021 [Technology Transfer Activities in Universities and Public Research Organizations: A Literature Overview](#). Ugo Finardi, Rolfo Secondo, Isabella Bianco.

2020

N. 12/2020 [Unexpected loss multiperiodale e pricing del rischio di credito](#). Franco Varetto.

N. 11/2020 [La ricerca in Nanotecnologie e Nanoscienze in Italia: spesa del settore pubblico e aree tematiche prevalenti](#). Ugo Finardi, Andrea Orazio Spinello.

N. 10/2020 [Persistent fast growth and profitability](#). Lucio Morettini, Bianca Potì, Roberto Gabriele.

N. 9/2020 [Binomio Burnout e Mindfulness nelle organizzazioni. Alcuni studi e scenari di applicazione](#). Oriana Ippoliti, Riccardo Briotti, Bianca Crocamo, Antonio Minopoli.

N. 8/2020 [Innovation and communication of companies on Twitter before and during COVID-19 crisis](#). José N. Franco-Riquelme, Antonio Zinilli, Joaquín B. Ordieres-Meré and Emanuela Reale.

N. 7/2020 [The proposal of a new hybrid methodology for the impact assessment of energy efficiency interventions. An exploratory study](#). Monica Cariola, Greta Falavigna.

N. 6/2020 [The technology innovative system of the Silicon Valley](#). Angelo Bonomi.

N. 5/2020 [Storia dell'industria delle macchine utensili in Piemonte dalle origini alla seconda guerra mondiale](#). Secondo Rolfo.

- N. 4/2020 [Blockchain e Internet of Things per la logistica Un caso di collaborazione tra ricerca e impresa](#). Edoardo Lorenzetti, Lucio Morettini, Franco Mazzenga, Alessandro Vizzarri, Romeo Giuliano, Paolo Peruzzi, Cristiano Di Giovanni.
- N. 3/2020 [L'impatto economico e fiscale di un evento culturale: misure e scala territoriale](#). Giovanna Segre, Andrea Morelli.
- N. 2/2020 [Mapping the tangible and intangible elements of the historical buildings and spaces](#). Edoardo Lorenzetti, Nicola Maiellaro.
- N. 1/2020 [Il lavoro agile negli enti pubblici di ricerca](#). Emanuela Reale, Serena Fabrizio, Andrea Orazio Spinello.
- 2019
- N. 6/2019 [Women's candidatures in local elections: does the context matter? Empirical evidence from Italian municipalities](#). Igor Benati, Greta Falavigna, Lisa Sella.
- N. 5/2019 [Research activities in Nanotechnologies and Nanosciences: an analysis of Piedmont's nanotech research system](#). Ugo Finardi.
- N. 4/2019 [Xylella fastidiosa: patogenesi, danni economici e lotta al disseccamento rapido dell'olivo](#). Maurizio Conti.
- N. 3/2019 [Flussi di traffico attraverso il tunnel automobilistico del Frejus: un semplice esercizio di forecasting e alcune considerazioni a margine](#). Ugo Finardi.
- N. 2/2019 [The Start-up Venture Capital Innovation System Comparison with industrially financed R&D projects system](#). Angelo Bonomi.
- N. 1/2019 [Complessità delle organizzazioni, complessità della formazione. Report di studio qualitativo ed analisi ermeneutica del Modello TRASE – IRCRES/CNR-IMO](#). Anna Chiara Scardicchio.
- 2018
- N. 13/2018 [Competenze di sviluppo sistemico evolutivo per la leadership e le organizzazioni orizzontali](#). Erica Rizziato, Erika Nemmo.
- N. 12/2018 [Organizzazioni e leadership orizzontali: il percorso di training sistemico evolutivo \(TRASE\)](#). Erica Rizziato.
- N. 11/2018 [Point-in-time vs. through-the-cycle: filosofie di rating a confronto](#). Franco Varetto.
- N. 10/2018 [Evaluating social innovation: results and emerging issues from a random-trial evaluation of a program for the inclusion of migrant adolescents](#). Valentina Lamonica, Elena Ragazzi, Lisa Sella.
- N. 9/2018 [Promozione dell'Imprenditorialità nelle Nuove Tecnologie. Caso Studio: Associazione "La Storia nel Futuro"](#). Angelo Bonomi.
- N. 8/2018 [Nanotechnology patenting in Piedmont: analysis and links with research and industrial environment in the Region](#). Ugo Finardi.
- N. 7/2018 [I canali innovativi di industria 4.0 e le PMI](#). Angelo Bonomi.
- N. 6/2018 [Does the construction of biogas plants affect local property values?](#). Marco Modica.
- N. 5/2018 [Public research in Nanotechnology in Piedmont \(Italy\)](#). Ugo Finardi.
- N. 4/2018 [Le tecnologie di Industria 4.0 e le PMI](#). Angelo Bonomi.
- N. 3/2018 [Overcoming sustainability barriers through Formalized Network Contracts \(FNCs\): the experience of Italian SMEs](#). Laura Corazza, Maurizio Cisi, Greta Falavigna.
- N. 2/2018 [A new taxonomy of rewards in public administration for management implications](#). Mario Coccia, Igor Benati.
- N. 1/2018 [The role of collective remittances in community development: the case of Hometown Associations](#). Barbara Bonciani.
- 2017
- N. 14/2017 [La correlazione tra PD ed LGD nell'analisi del rischio di credito](#). Franco Varetto.
- N. 13/2017 [L'Alterità nel "Terzo Paesaggio": una risorsa di sostenibilità socio-territoriale. Gli stranieri residenti nella Valle del Turano, fonte di rinascita tra innovazione e tradizione. Presentazione progetto di ricerca etnografica](#). Brunella Bonetti.
- N. 12/2017 [Long time series of highly cited articles: an empirical study](#). Ugo Finardi.
- N. 11/2017 [Un'analisi multidimensionale della sostenibilità per l'agricoltura familiare. Il caso dell'area amazzonica peruviana](#). Andrea Pronti, Flavio Bertinaria.
- N. 10/2017 [Cosa s'intende per sostenibilità economica? Riflessione sul significato di sistema economicamente sostenibile](#). Giuseppe Cornelli.
- N. 9/2017 [Supportare la ricerca e l'innovazione in sanità tramite i modelli organizzativi: il caso dell'Azienda Ospedaliera "SS. Antonio e Biagio e Cesare Arrigo" di Alessandria](#). Greta Falavigna,

In the framework of the RAISE project, Spoke 5 - P6 Education, the online community STACY (Secondary TeAcher CommuniTY) was created in 2023, comprising teachers and researchers with the aim of fostering a productive exchange between the realms of research and education concerning the understanding and advancement of Artificial Intelligence (AI), Robotics, and Anticipatory Thinking. Within this framework, a study has been initiated to undertake metalinguistic analysis of the terminology used in the communication and portrayal of AI and Robotics in the Italian mainstream media. The aim is to enhance and strengthen linguistic resources conducive to effective shared communication within the specialized language of AI and Robotics. To this end, a metalinguistic reflection activity is proposed on a series of terms commonly utilized in the standard Italian language, related to the semantic field of AI and Robotics. This selection will be based on frequency analysis derived from a review of seven widely circulated Italian newspapers (La Repubblica, Corriere della sera, Il Secolo XIX, Il Foglio, Il Fatto quotidiano, Corriere del Mezzogiorno, Libero). The findings will be presented in the form of a thesaurus, including a glossary of sector-specific terms, delineated across two distinct registers: the common language prevalent in generalist newspapers and magazines, and the specialized.