L'uso delle *keyword* nella comunicazione scientifica, con un'analisi sul settore scientifico disciplinare M-STO/08 (HIST-04/C)

di Rossana Morriello, Lucia Sardo

1. Premessa – definizione e ambito di applicazione

Keyword o parola chiave, in italiano, indica, secondo la definizione del vocabolario Treccani

> con riferimento a un archivio di dati [...], ciascuno di quei dati che, in base alla loro posizione e al loro contenuto, permettono un rapido ed efficiente accesso alle informazioni o sono utilizzati per attribuire a queste un ordine logico (per es., in un file che contiene gli indirizzi, inseriti in ordine sparso, di un insieme di persone, la città può essere usata come parola chiave per guidare la ricerca delle informazioni, il cognome può essere usato come parola chiave per ordinare alfabeticamente l'insieme degli indirizzi); con questi sign., la locuz. è un calco dell'ingl. *keyword*[1].

La definizione data da Treccani non esplicita chiaramente la funzione delle *keyword* nei diversi contesti in cui possono trovare impiego, soprattutto online; la definizione dell'Oxford English Dictionary chiarisce meglio: «A word (usually one of several) chosen to indicate or represent the content of a larger document, text, record, etc., in an index, catalogue, or database. Later also: any word entered as a search term in a database or search engine»[2]. In questo senso il termine appare per la prima volta nel 1827[3].

ROSSANA MORRIELLO, Università degli studi di Firenze, e-mail: rossana.morriello@unifi.it
LUCIA SARDO, AMS Università di Bologna, lucia.sardo@unibo.it
Ultima consultazione siti web: 30 giugno 2025.

* Le autrici hanno condiviso l'intero contenuto dell'articolo ma nello specifico i paragrafi 1, 2 e 6 vanno attribuiti a Lucia Sardo; i paragrafi 3, 4 e 5 vanno attribuiti a Rossana Morriello.

1 <https://www.treccani.it/vocabolario/parola/>.

2 Oxford English Dictionary. *Keyword*, Revised 2012, <https://www.oed.com/dictionary/*keyword*_n?tab=meaning_and_use>.

3 L'Oxford English Dictionary riporta così i dettagli della citazione: «1827: A quantity of a writer's text is thrust into the page, far beyond the point at which all connection with the key-word fully terminates.

AIB studi, vol. 65 n. 1 (gennaio/aprile 2025), p. 67-84. DOI 10.2426/aibstudi-14149
ISSN: 2280-9112, E-ISSN: 2239-6152 - Copyright © 2025 Rossana Morriello, Lucia Sardo

Online si trovano molte definizioni diverse di *keyword*, generalmente in siti di editori scientifici (un esempio fra i molti, «le parole chiave sono parole o frasi che rappresentano gli argomenti o le idee principali presentate in un documento di ricerca»[4]), non sempre univoche, ma in linea di massima le principali di esse si concentrano sul fatto che sono parole utili a reperire documenti, in particolar modo nell'ecosistema digitale. Sono definizioni però limitanti e soprattutto, nell'accezione con cui le *keyword* vengono usate nell'editoria scientifica, non corrette, dato che, ad esempio, spesso una *keyword* può essere costituita da più termini, e che la loro finalità non è il semplice reperimento di documenti. Lo standard americano *JATS: Journal Article Tag Suite* definisce così:

> <kwd> *Keyword***.** A *keyword* is a finding aid used to record one subject term, key phrase, abbreviation, indexing word, taxonomic structure, or other expression that is associated with a whole structure (such as an article). *Keyword*s can be used for identification, searching, and indexing purposes. A *keyword* element (<kwd>) may contain a repetition of words found in the structure, or it may contain broader terms, narrower terms, related terms, or taxonomic identifiers used to find the structure in a search or provide semantic classification of the text by mapping a specific structure to a taxonomy[5].

L'International encyclopedia of information and library science definisce così sinteticamente *keyword*:

> a word that succinctly and accurately describes the subject, or an aspect of the subject, discussed in a document. In a permuted title index, the word from the title of a document considered to be most indicative of the subject matter, i.e. the *keyword*, can be used as an indexing term, permuted by one of a number of methods: KWIC (KeyWord in context), KWAC (KeyWord and context) or KWOC (KeyWord out of context)[6].

La parola *keyword* assume significati diversi in diversi contesti, ma progressivamente, nell'uso comune, il suo significato è riassumibile con termini o frasi che descri-

Monthly review April 350»; la seconda attestazione è del 1864, con la seguente citazione: «1864: The first words of each quotation are given in alphabetical order, but the same passage is also given under what he considers to be the key-word. *C. T. Ramage, Beautiful Thoughts from Latin Authors p. ix*». Cfr. <https://www.oed.com/dictionary/*keyword*_n?tab=meaning_and_use>.

4 Gilberto De Abreu, *Parole chiave in un documento di ricerca: L' importanza della scelta giusta*, «Mind the graph», 14 giuno 2023, <https://mindthegraph.com/blog/it/parole-chiave-in-un-documento-di-ricerca/>.

5 *ANSI/NISO Z39.96-2019, JATS: Journal article tag suite, version 1.2*, <https://www.niso.org/publications/z3996-2019-jats>. Lo standard ha la finalità di definire un insieme di elementi e attributi XML che descrivono il contenuto e i metadati degli articoli di riviste (articoli di ricerca e non, lettere, editoriali, recensioni di libri e prodotti) con l'intento di fornire un formato comune con cui gli editori e gli archivi possano scambiare i contenuti delle riviste. L'intento della Tag suite è quello di preservare il contenuto intellettuale delle riviste indipendentemente dalla forma in cui tale contenuto è stato originariamente fornito.

6 *International encyclopedia of information and library science*, edited by John Feather and Paul Sturges, 2. ed. London: Routledge, 2003, p. 341.

vono il contenuto di una risorsa e la cui finalità principale è l'ottimizzazione dei risultati per i motori di ricerca, soprattutto nel contesto del web marketing e della SEO (Search engine optimization). Nell'ambito più specifico dell'editoria scientifica e accademica, le *keyword* invece sono termini o frasi brevi che riassumono i concetti principali trattati in un articolo; le finalità per cui vengono redatte sono molteplici e verranno analizzate più avanti[7].

Da un punto di vista storico, per la lingua inglese, la parola *keyword* ha iniziato ad assumere grande importanza dalla fine degli anni Cinquanta del secolo scorso, in due contesti e con due significati diversi; come afferma Rosenberg:

> beginning about 1958, *keyword* mattered in English, but, from the start, it mattered in two different ways and in two very different areas: literary and cultural theory, on the one hand, and computer and information theory, on the other. Yet, during the late twentieth century, these two senses came eventually not only to interact but also to reinforce and invigorate each other in ways that tell us a great deal about the emerging epistemological environment[8].

Nel contesto 'letterario' e culturale il significato di *keyword* è stato definito dal lavoro di Raymond Williams, in particolare in *Culture & society, 1780-1850*[9] e nel seguente *Keywords: a vocabulary of Culture & society*[10] del 1976. Invece il senso del termine nel contesto della teoria dell'informazione e dell'informatica è stato promosso dai lavori di Hans Peter Luhn, ingegnere dell'IBM, e innovatore in molti ambiti, in particolare in articoli pubblicati fra il 1958 e il 1962, fra cui *Keyword-in-context index for technical literature* del 1960[11],

Sempre Rosemberg sottolinea che

> evidently, the context of information retrieval is very different from that of literary criticism, but the contrast between the usages of Williams and Luhn is even stronger than the difference between their fields would suggest. When Luhn talked about *keywords*, he did not assume, as did Williams, that these ought to be either problematical or deep; by *keyword*, he meant any word that characterized a document because it occurred in it in statistically notable ways, or conversely any word provided in an information query used to search that same document. The *Keyword-in- context (KWIC)* index discussed in Luhn's 1960 article is essentially a machine- generated concordance[12].

7 In senso stretto, si dovrebbe distinguere fra *keyword*s (singole parole) e *keyphrares* (frasi brevi); per comodità e anche perché spesso i due termini non sono distinti, quando parliamo di *keyword* intendiamo sia le *keyword* in senso stretto che le *keyphrases*.

8 Daniel Rosenberg, *Keyword*. In: *Information*. *Keywords*, edited by Michele Kennerly, Samuel Frederick, and Jonathan E. Abel. New York: Columbia university press, 2021, p. 122-123.

9 Raymond Williams, *Culture and society, 1780– 1850*. London: Chatto & Windus, 1958.

10 Raymond Williams, *Keywords: A vocabulary of Culture and society.* New York: Oxford university press, 1976.

11 Hans Peter Luhn, *Keyword-in-context index for technical literature*, «American documentation», 11 (1960), n. 4, p. 288-295.

12 D. Rosenberg, *Keyword* cit., p.123.

L'idea delle KWIC, in ambito biblioteconomico, può essere fatta risalire alle teorizzazioni di Andrea Crestadoro che nel suo *The art of making catalogues*[13] del 1856 accenna a questo metodo – non esplicitamente e non usando questa dicitura per fare riferimento al sistema che propone – per facilitare l'accesso al posseduto di una biblioteca, e lo mette in pratica nei cataloghi che redige per la biblioteca di Manchester[14].

Senza addentrarsi in una analisi del valore che possono avere le *keyword* e degli aspetti problematici legati al loro uso[15], la situazione attuale risulta ovviamente diversa da quella degli anni Sessanta del Novecento, sia in ambito editoriale che informatico. Va ricordato che il loro uso in ambito accademico e di editoria scientifica inizia a diffondersi negli anni Settanta del secolo scorso, soprattutto nelle discipline che afferiscono al settore delle scienze dure e della medicina, per poi progressivamente diffondersi in tutti gli altri settori dagli anni Ottanta in poi. È dagli anni Novanta che le riviste iniziano a chiedere formalmente le *keyword* agli autori degli articoli, in aggiunta all'abstract, già presente da decenni.

Nello specifico di nostro interesse, il lavoro di Landera e Hjorland[16] è fondamentale per chiarire il significato di *keyword* e di termini correlati o correlabili. L'articolo offre una esaustiva disamina dei significati della parola *keyword* e di termini correlati come '*keyphrase*', '*descriptor*', '*index term*', '*subject heading*', '*tag*' and '*n-gram*'.

In sintesi, per *keyword* gli autori propongono la seguente definizione:

> A word, selected for its significance for describing the content of a document, whether or not the selection is done by man or machine, whether it is a controlled or uncontrolled term, or whether it is derived from or assigned to a document.
> A *keyword* may be derived from a subject heading, but *subject heading* and *descriptor* are terms related to specific kinds of indexing languages.
> A *keyword* is different from a full multiterm subject heading or index term, which is called a *keyphrase*. Both *keyword*s and keyphrases are sometimes included in the term *key terms*[17].

In questo caso è interessante rilevare che gli autori indicano anche cosa non sono le *keyword* (e, aggiungo, di conseguenza, come non andrebbero usate). Non sono, ad esempio, stringhe di soggetto, per cui l'individuazione di parole chiave non corrisponde al risultato di una analisi semantica di un documento 'riassunta' – appunto – in una stringa di soggetto, perché così facendo si rischia di non rendere effettivamente ed efficacemente il contenuto del documento.

13 Andrea Crestadoro, *The art of making catalogues of libraries; or a method to obtain in a short time a most perfect, complete, and satisfactory printed catalogue of the British Museum Library*. London: Literary, scientific & artistic reference office, 1856.

14 Cfr. quantomeno Muhammad Riaz, *Advanced indexing and abstracting practices*. New Delhi: Atlantic publishers & distributors, 1989, p. 41 e seg.; Alfredo Serrai, St*oria della bibliografia. Vol. X: Specializzazione e pragmatismo: i nuovi cardini della attività bibliografica*. Roma: Bulzoni, 1999, p. 971 e seguenti.

15 Per un approfondimento del tema si rimanda quantomeno a Betsy Van der Veer Martens, *Keywords in and out of context*. Switzerland: Springer, 2023.

16 Marco Lardera; Birger Hjorland. *Keyword*, «Knowledge organization», 48 (2021), n. 6, p. 430-456. DOI: 10.5771/0943-7444-2021-6-430.

17 *Ivi*, p. 437.

La disamina linguistica sui possibili significati e sulle possibili interpretazioni del termine non sono oziose, in quanto il loro uso e la loro realizzazione dipendono inevitabilmente dal significato che gli autori assegnano alle *keyword* e ciò condiziona a sua volta la scelta delle parole usate.

Se, quindi, è possibile trovare valide definizioni di *keyword*, sia per l'ambito più generale della ricerca online, sia per quello più specifico della ricerca e dell'editoria scientifica e accademica, mancano del tutto delle analisi sul loro effettivo uso da parte, quantomeno, della comunità scientifica che si occupa di biblioteconomia e bibliografia (in senso ampio) e discipline affini[18] e mancano anche linee guida o direttive di qualsiasi sorta per aiutare nella creazione di *keyword* che siano efficaci e utili in ambito scientifico accademico.

2. *Keyword* nell'editoria scientifica e accademica: cosa sono e a cosa servono

Come accennato sopra, l'uso delle *keyword* negli articoli scientifici è una pratica che si è diffusa a partire dagli anni Settanta-Ottanta del secolo scorso, con l'avvento delle banche dati quali PubMed o Web of Science e con l'aumento dell'importanza delle basi dati bibliografiche e dell'indicizzazione automatica (con i precedenti legati all'elaborazione dei sistemi di indicizzazione KWIC a partire dalla fine degli anni Cinquanta) e che si diffonde nell'ultimo decennio del secolo scorso.

Le *keyword* diventano sempre più importanti per migliorare la ricercabilità degli articoli (in un mondo dove le pubblicazioni scientifiche aumentano globalmente in modo esponenziale) aiutando i motori di ricerca sia generalisti che accademici (ad esempio, Google Scholar) nell'indicizzazione, e dovrebbero permettere una migliore classificazione tematica da parte delle banche dati (banche dati che spesso prevedono anche l'inserimento di altre parole chiave oltre a quelle indicate dall'autore per aumentare il grado di reperibilità delle risorse e la creazione di cluster tematici). Inoltre, fattore questo da non sottovalutare, vengono spesso usate anche per suggerire revisori con competenze nei rispettivi ambiti di studio e ricerca.

Esse hanno una grande rilevanza non solo per accedere alle pubblicazioni (tralasciamo i problemi linguistici che sorgono una volta individuati i documenti grazie ad abstract e *keyword* in inglese ma che sono scritti in lingue non conosciute) ma anche per categorizzare, dare visibilità alle pubblicazioni stesse, individuare esperti, consentire match alle macchine, fattore questo sempre più utilizzato e diffuso.

Per quanto riguarda invece la creazione delle *keyword*, solo editori o altri player commerciali danno indicazioni su come farlo[19]. In questo senso, quanto suggerito da Spinger in una sezione dedicata al supporto agli autori sulla creazione di abstract e *keyword* è accurato e preciso; sinteticamente l'editore ricorda che sono uno strumento fondamentale per aiutare indicizzatori e motori di ricerca ad individuare articoli rilevanti, e la loro finalità è evidente: «If database search engines can find your

18 Possiamo dire della comunità che afferisce al settore scientifico disciplinare ex M-STO/08 ora HIST-04/C, Archivistica, bibliografia e biblioteconomia.

19 Alcuni esempi fra molti: Elizabeth George, *Identifying keywords for scientific papers: a simple 5-step guide*. «Researcher.life», 20 February 2023, <https://researcher.life/blog/article/identifying-*keyword*s-for-scientific-papers-a-simple-5-step-guide/>; *Title, abstract and keywords*, <https://www.springer.com/it/authors-editors/authorandreviewertutorials/writing-a-journal-manuscript/title-abstract-and-*keyword*s/10285522?srsltid=AfmBOopFzSPzbG2sKMbTvGGL-OsYXwcjfdxTN64_Qosyr7SIU_9iyvrG>; *How to choose keywords for a manuscript?* <https://scientific-publishing.webshop.elsevier.com/manuscript-preparation/how-choose-*keyword*s-manuscript/>.

journal manuscript, readers will be able to find it too. This will increase the number of people reading your manuscript, and likely lead to more citations»[20]. Sempre Springer ricorda che per essere efficaci le *keyword* devono non solo essere scelte con cura ma devono rappresentare il contenuto dell'articolo ed essere specifiche rispetto al proprio ambito o sotto-ambito di studio. I suggerimenti dati sono i seguenti:
1. Include your techniques and/or specific methodology.
2. Focus on the main topic of your research.
3. Avoid *keyword*s that are only one word.
4. Avoid overlapping *keyword*s in your title and those in your *keyword* list.
5. Follow the journal guidelines when selecting *keyword*s.
6. Perform *keyword* research before submitting your article.

Esistono infine strumenti specifici per la creazione di *keyword*, pensati principalmente per il mondo del web, ma potenzialmente utili anche per il mondo della ricerca, quali *keyword planner* di Google[21] e manuali dedicati alla stesura di articoli scientifici, tesi e dissertazioni[22].

3. Le *keyword* come strumento di disseminazione e comunicazione scientifica

Come nelle indicazioni dell'editore Springer sopra citate, le parole chiave sono utili in generale per aumentare il numero di persone che leggono un articolo e dunque incrementare le citazioni. La correlazione tra l'efficacia delle *keyword* e il numero di citazioni è comprovata da alcuni studi bibliometrici[23]. Dal punto di vista di un editore si tratta, quindi, di un aspetto centrale poiché è legato all'impatto della rivista e comprende dei risvolti economici che si traducono in interesse a sottoscrivere gli abbonamenti da parte delle biblioteche o altre istituzioni. L'impatto di una rivista viene oggi misurato, in molti casi, attraverso la lente degli indici bibliometrici e si genera così un circolo, virtuoso o vizioso a seconda dei punti di vista, per cui quanto più un articolo viene letto e riceve citazioni, tanto più alti sono gli indicatori bibliometrici della rivista, quali l'impact factor. Per tale ragione gli editori scientifici dedicano molta attenzione all'indicizzazione dei contenuti delle banche dati editoriali e all'uso delle *keyword*.

Dal punto di vista dell'autore, il processo è analogo poiché gli indici bibliometrici vengono utilizzati nella valutazione della ricerca e quindi per valutare i singoli ricercatori, i dipartimenti, gli atenei. La valutazione bibliometrica riguarda principalmente i settori STEM (Science Technology Engineering Mathematics) ma ormai indici come l'*impact factor* oppure il conteggio delle citazioni vengono considerati anche per le discipline umanistiche. Alcune riviste italiane di biblioteconomia sono presenti nelle principali banche dati citazionali, Scopus e Web of Science, e dunque tali dinamiche riguardano il settore scientifico disciplinare HIST-04/C Archivistica, bibliografia e biblioteconomia. Tuttavia, tralasciando il complesso mondo della bibliometria e della valutazione della ricerca, rimane il fatto che le parole chiave svolgono un ruolo fondamentale nella trasmissione del contenuto di un articolo

20 Cfr. *Abstract and keywords*, <https://www.springer.com/gp/authors-editors/journal-author/abstract-and-*keyword*s/1380>.

21 Lo strumento è disponibile al seguente link: <https://business.google.com/in/ad-tools/*keyword*-planner/>.

22 Kate L. Turabian, *A manual for writers of research papers, theses, and dissertations*, 9ª ed. Chicago: The University of Chicago press, 2018.

scientifico ai lettori, tanto più quando l'articolo è in open access. In questo caso, quando parliamo di lettori non ci riferiamo solo ai pari, ovvero ricercatori, docenti, esperti della materia, ma anche potenzialmente a qualsiasi cittadino. La lenta ma progressiva affermazione del movimento open access da un lato[24] e l'attenzione crescente alla terza missione delle università dall'altro, sono i fattori che modificano parzialmente la natura e gli obiettivi della comunicazione scientifica. L'open access implica che le pubblicazioni siano liberamente accessibili a tutti e dunque gli articoli ad accesso aperto sono ricercabili da qualsiasi utente nel web, esperto o non esperto della materia. Al contrario, di quanto avviene per le risorse chiuse dietro a un paywall, consultabili dai docenti e dai ricercatori afferenti alle istituzioni universitarie grazie agli abbonamenti e alle licenze d'uso sottoscritte dalle biblioteche. La terza missione viene definita dall'ANVUR come «la propensione delle strutture all'apertura verso il contesto socio-economico, esercitato mediante la valorizzazione e il trasferimento delle conoscenze»[25]. Il presupposto è dunque proprio l'apertura degli atenei e dei risultati della ricerca accademica, incluse le pubblicazioni, al territorio e dunque a un pubblico generico. L'idea alla base della terza missione è che l'università debba restituire i risultati degli investimenti pubblici alla collettività, ed è ampiamente promossa anche a livello internazionale nei bandi per il finanziamento dei progetti di ricerca, che ormai in gran parte richiedono che gli output della ricerca finanziata siano resi disponibili in open access. La maggiore apertura indotta dai due fattori sopra indicati, amplia il pubblico a cui sono destinati i processi di disseminazione e comunicazione scientifica nelle università, includendovi potenzialmente chiunque, dal giornalista al semplice curioso, e non solo i pari o gli esperti della materia.

Inoltre, con l'evoluzione dell'ecosistema digitale i canali di scoperta e recupero dell'informazione scientifica si sono moltiplicati e includono gli OPAC o i discovery tool delle biblioteche, le banche dati degli editori, i repository istituzionali e disciplinari aperti. L'accesso avviene direttamente tramite questi canali oppure con l'intermediazione di un motore di ricerca nel web, come Google. Le *keyword* sono funzionali a qualsiasi sistema di information retrieval e sono centrali per la *discoverability* sia nelle banche dati sia nel web dove, come prima funzione, contribuiscono a determinare il ranking dei risultati. Tale centralità diviene cruciale nell'attuale sistema della comunicazione scientifica caratterizzato da una quantità di pubblicazioni in

23 Shahadat Uddin; Arif Khan, *The impact of author-selected keywords on citation counts*, «Journal of infometrics», 10 (2016), n. 4, p. 1166-1177, DOI: 10.1016/j.joi.2016.10.004.

24 Per i dati sullo sviluppo del movimento open access si vedano i contributi: Rossana Morriello, *Open access: verso una conoscenza aperta e condivisa*, e Nino Grizzuti, Francesca Rossi, Antonio Scolari, *I contratti trasformativi*. In: *Rapporto sulle biblioteche italiane 2021-2023*, a cura di Vittorio Ponzani e Maddalena Battaggia; presentazione di Laura Ballestra. Roma: AIB, 2024.

25 Secondo la definizione data da ANVUR (Agenzia nazionale di valutazione del sistema universitario e della ricerca) fin dal primo esercizio di valutazione della ricerca (VQR) la terza missione è costituita dal «contributo delle Istituzioni allo sviluppo della società e del territorio, favorendo il dialogo tra Istituzioni, le imprese e la società civile», cfr. ANVUR, *Valutazione della qualità della ricerca 2004 2010 (VQR 2004 2010)*, 7 novembre 2011, p. 15, <https://www.anvur.it/sites/default/files/2025-02/bando_vqr_def_07_11.pdf>. L'articolazione della terza missione è andata perfezionandosi e ampliandosi nel corso del tempo per l'agenzia di valutazione, fino ad includere le svariate attività indicate per l'ultima VQR 2020-2024 <https://www.anvur.it/it/ricerca/valorizzazione-delle-conoscenze>.

crescita costante, in parte causata dalla cultura del *publish or perish*, per cui si parla ormai di *discoverability crisis*. In altre parole, il recupero dell'informazione scientifica pertinente diventa sempre più difficile e di conseguenza le strategie di SEO sono fondamentali.

Per quanto riguarda il mondo accademico si rilevano delle specificità che hanno portato a coniare la definizione di ASEO (Academic search engine optimization). Nell'ASEO sono tre gli elementi che favoriscono l'indicizzazione efficace e aumentano la possibilità per l'utente di recuperare l'informazione ricercata: il titolo, l'abstract, e le *keyword*. Nelle banche dati full-text si aggiunge ovviamente il testo pieno delle risorse, ma la presenza di titolo, abstract e soprattutto di parole chiave consente di restituire risultati più precisi nella ricerca, riducendo il rumore. Le funzioni delle *keyword* nell'information retrieval, se viste dal lato utente, sono molteplici. Innanzitutto, permettono di collocare l'articolo in una linea di ricerca specifica. Questo consente agli studiosi di capire subito il tema dell'articolo trovato e di appurare se è d'interesse per il proprio quesito di ricerca, e quindi di recuperare la bibliografia pertinente con maggiore precisione e rapidità. Nel caso delle pubblicazioni ad accesso aperto depositate nelle banche dati, nei repository istituzionali, nei repository di *preprint*, gli articoli sono accessibili a tutti, anche al lettore non esperto e le parole chiave rappresentano un ausilio ma introducono anche una certa complessità perché il lettore esperto e il lettore generico necessitano di approcci diversi nella definizione delle parole chiave per descrivere il contenuto di una pubblicazione.

Jean-Luc Lebrun per analizzare e affrontare questa complessità individua tre tipi di *keyword*: generali, intermedie e specifiche, raffigurandole in un triangolo invertito (Figura 1). Nel triangolo le *keyword* specifiche si trovano nella punta collocata in basso e rappresentano la massima profondità. Nella cima della piramide rovesciata vi sono le *keyword* generali che rappresentano la massima ampiezza. Ampiezza e profondità sono inversamente proporzionali, maggiore l'ampiezza delle parole chiave, minore la profondità.

L'ampiezza e la profondità non sono qualità intrinseche ma dipendono dalla frequenza con la quale vengono usate all'interno del canale di disseminazione di un determinato articolo[26]. Per esempio, una rivista di biblioteconomia può considerare la parola 'biblioteca' molto generica mentre per una rivista di cultura generale, o comunque di altro ambito, può essere una parola specifica. Un elemento di contesto da considerare è dunque la tipologia di 'contenitore' (rivista, repository, banca dati) e la sua specializzazione. Un altro elemento importante, correlato al precedente, è la tipologia di lettore a cui la rivista si rivolge: per una rivista indirizzata a un pubblico di non esperti del settore, anche una *keyword* generica può diventare specifica.

Nella definizione delle parole chiave per un articolo occorrerebbe dunque mettersi nei panni del lettore al quale ci si rivolgere e pensare come tale lettore potrebbe cercare la risorsa. Tuttavia, vi sono molte variabili da considerare. In primo luogo, la tipologia di rivista e il suo pubblico in relazione alla disponibilità potenziale, anche futura dell'articolo. Per esempio, come già accennato, il pubblico potenziale è diverso se la rivista è open access oppure dietro *paywall*. Nel primo caso, indipendentemente dal taglio e dalla specializzazione della rivista, i contenuti saranno comunque accessibili a tutti. Ma anche una rivista chiusa potrebbe diventare open access in futuro oppure essere o diventare ibrida. Nel caso di deposito dell'articolo, o *preprint*, in un repository ad accesso aperto è analogamente diverso il caso di un repository disciplinare, dunque specializzato, e di un repository istituzionale, per esempio di un ateneo, tipicamente multidisciplinare.

In base a tali considerazioni l'autore dovrebbe quindi mirare a un bilanciamento ideale tra ampiezza e profondità. Come suggerisce Lebrun, sarebbe opportuno attribuire sempre al proprio articolo *keyword* a più di un livello del triangolo.

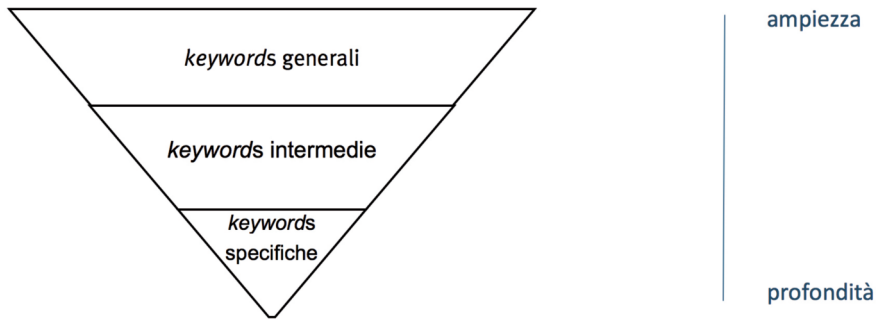

Figura 1 – Piramide rovesciata delle *keyword*, adattamento da Lebrun, 2007.

Il livello di profondità si determina anche con l'uso di un linguaggio appropriato. Soprattutto quando si rivolgono a un pubblico generico o potenzialmente generico, gli autori dovrebbero utilizzare un linguaggio comune ed evitare di creare terminologia nuova, di usare abbreviazioni e termini molto tecnici oppure slang scientifico. È stato rilevato, per esempio, come l'uso di termini nuovi attribuiti dagli autori ai propri articoli nelle *keyword* abbia una correlazione negativa con il numero delle citazioni ricevute[27]. In particolare, quando si fa riferimento a concetti innovativi, questi dovrebbero essere verificati, come suggerisce la norma ISO 5963:1985 *Documentation - methods for examining documents determining their subjects and selecting indexing terms* (norma tuttavia in fase di revisione[28]), per individuare la loro forma accettata, in dizionari ed enciclopedie specializzate nel settore, thesauri, schemi di classificazione[29]. Altri metodi sono la consultazione di esperti o pari oppure la verifica delle *keyword* usate in articoli di argomento simile.

Prendere a riferimento un vocabolario controllato, come quello di un thesaurus, aiuta ovviamente nella standardizzazione necessaria all'information retrieval, ma l'uso non può essere lasciato al singolo autore senza supporto. Alcune riviste prevedono l'utilizzo di soggettari come il «Journal of information science» che in fase di bozza dell'articolo attribuisce automaticamente le *keyword* derivate dalla banca dati Library and Information Science Abstract (LISA)[30], il «Journal of economic literatu-

26 Jean-Luc Lebrun, *Scientific writing. A reader and writer's guide*. Singapore: World scientific, 2007, p. 111.

27 S. Uddin; A. Khan, *The impact of author-selected keywords on citation counts* cit.

28 La norma sarà sostituita da ISO/AWI 5963, attualmente in fase di sviluppo.

29 Aditi Roy; Saptarshi Ghosh, *Freedom versus standard in article keyword generation: an empirical study*, «Journal of library metadata», 24 (2024), n. 4, p. 292-293.

30 James Hartley; Ronald N. Kostoff, *How useful are 'key words' in scientific journals?*, «Journal of information science», 29 (2003), n. 5, p. 433-438, DOI: 10.1177/0165551030295008.

re»[31], utilizzato dalla banca dati Econlit, oppure il Medical Subject Headings (MeSH), utilizzato nella banca dati Medline, i cui articoli però vengono indicizzati dai bibliotecari della National Library of Medicine[32], con l'obiettivo esplicito di facilitare l'online information retrieval in particolare in PubMed[33].

Come già evidenziato, l'uso libero di parole chiave da parte degli autori può avere dei limiti sia legati all'attitudine dell'autore, sia collegati alla discrepanza tra ampiezza e intensità che si genera in certe tipologie di articoli. Tuttavia, anche i vocabolari controllati possono creare dei problemi per l'aggiornamento non costante. Per esempio, termini come '*scientometrics*' oppure '*altmetrics*', '*big data*' o '*blockchain*' che si possono trovare in una rivista di biblioteconomia, non sono presenti in tutti i vocabolari controllati e thesauri[34].

Le banche dati bibliografiche, full-text o citazionali come Scopus e Web of Science, usano sistemi di indicizzazione semantica automatizzati, come vedremo meglio più avanti, ma in tutti questi casi si tratta di operazioni controllate centralmente all'interno di banche dati altamente strutturate e continuamente aggiornate. In assenza di strutture di questo genere, la soggettazione lasciata al singolo autore dell'articolo non risulta in generale soddisfacente. Come si scriveva sopra, stringhe di soggetto e *keyword* sono due cose diverse, e le prime, oltre ad essere sistemi complessi, risultano non efficaci in molti casi, in particolare per le discipline umanistiche, per la varietà e specificità dei contenuti, la diversità linguistica, la variazione dei significati nel tempo, la difficoltà dei sistemi di indicizzazione di cogliere le sfumature di significato e la loro trasformazione nel tempo[35].

Le considerazioni collegate all'attribuzione di parole chiave ai fini del recupero dell'informazione assumono maggiore valenza alla luce del crescente utilizzo di sistemi di intelligenza artificiale nella comunicazione scientifica, sia da parte degli editori, sia da parte delle aziende di IA che si indirizzano ai contenuti scientifici in quanto affidabili, organizzati e indicizzati, e dunque dataset affidabili per il training dei modelli. Gli editori si stanno infatti avviando verso il nuovo business della vendita dei dati alle aziende di IA, generando proteste per la poca trasparenza e perché a volte lo fanno senza il consenso degli autori[36].

31 «Journal of economic literature (JEL), <https://www.aeaweb.org/econlit/jelCodes.php>.

32 Niki Baumann, *How to use the medical subject headings (MeSH),* «International journal of clinical practice», 70 (2026), p. 171-174, DOI: 10.1111/ijcp.12767.

33 National Library of Medicine, *Use of MeSH in online retrieval*, 2023, <https://www.nlm.nih.gov/mesh/intro_retrieval.html>.

34 A. Roy; S. Ghosh, *Freedom versus standard in article keyword generation* cit., p. 302. L'analisi è stata condotta sulla rivista indiana di biblioteconomia «Annals of library and Information Studies» (ALIS), confrontata con sei vocabolari controllati: UNESCO thesaurus, EuroVoc Thesaurus, Humanities and social science thesaurus (HASSET), ECLAS thesaurus, ERIC (Education resources information centre) thesaurus, STW Thesaurus for economics. L'analisi più ampia oggetto dello studio ha incluso anche i soggetti LCSH per rilevare il livello di sovrapposizione con le parole chiave degli autori, risultato basso.

35 Koraljka Golub [*et al.*], *Subject indexing in humanities: a comparison between a local university repository and an international bibliographic service*, «Journal of documentation», 76 (2020), n. 6, p. 1193-1214, DOI: 10.1108/JD-12-2019-0231. Si veda anche Rebecca Green, *The role of relational structures in indexing for the humanities,* «Information services & use», 17 (1997), n. 2-3, p. 85-100, DOI: 10.3233/ISU-1997-172-303.

36 Rossana Morriello, *IA ed editoria scientifica: come la Direttiva UE sta cambiando le regole del gioco*, «Informatica umanistica e cultura digitale: il blog dell'AIUCD», 31 dicembre 2024, DOI: 10.58079/1301b.

Un'altra funzione utilizzata in misura crescente dalle riviste, proprio grazie alla disponibilità di applicazioni di intelligenza artificiale, è l'assegnazione di un articolo ai revisori per la peer review sulla base delle *keyword* dell'articolo. Diverse piattaforme editoriali e riviste utilizzano le parole chiave fornite dagli autori per identificare i revisori esperti negli argomenti trattati dall'articolo. In questo utilizzo, termini precisi facilitano l'individuazione di revisori competenti[37]. Tale modalità di matching tra articolo e revisori viene attuata anche per i progetti di ricerca e in tutti i casi in cui sia necessaria un'attività di peer review.

La centralità delle *keyword* nel contesto della ricerca scientifica, ormai divenuta internazionale e aperta al pubblico generico e la cui dimensione è amplificata dagli strumenti digitali e dall'intelligenza artificiale, rende indispensabile porvi una grande attenzione, che potrebbe cominciare con la revisione delle parole chiave, così come degli abstract, da parte dei revisori nei processi di peer review oppure dei board delle riviste.

4. Il caso di studio italiano per il settore M-STO/08

Per poter verificare l'uso delle *keyword* da parte del mondo della ricerca, abbiamo deciso di prendere in considerazione i prodotti della ricerca del settore scientifico disciplinare M-STO/08 presenti nei repository istituzionali degli atenei italiani che utilizzano il Current research information system (CRIS) IRIS, prodotto e distribuito dal Cineca. Si tratta di 39 atenei nei quali è presente il SSD[38] in questione su 67 atenei che utilizzano IRIS[39]. Il range temporale dell'analisi copre gli anni dal 2010 al 2023, motivo per cui è stata mantenuta la 'vecchia' dicitura del settore scientifico disciplinare, in quanto la riforma degli stessi è del 2024[40]. Le tipologie IRIS considerate sono (in ordine alfabetico): Altro, Articolo in rivista, Banca dati, Breve introduzione, Brevetto, Concordanza, Contributo in Atti di convegno, Contributo in volume (Capitolo o Saggio), Curatela, Edizione critica, Esposizione. Si tratta di 6805 record, di cui il 71,9% ha il campo parola chiave in una qualsiasi lingua compilato, con un numero di *keyword*s variabile (si veda infra).

Prima di presentare i risultati, è utile premettere che, com'è ovvio, gli autori non inseriscono *keyword* generiche o poco precise volontariamente poiché non è intenzione di nessuno e nessuna fornire informazioni che non rappresentino adeguatamente il contenuto dell'articolo. L'interesse di un autore è far conoscere i risultati della propria ricerca ma a volte la creazione di una scheda completa di tutti i metadati all'interno di un repository istituzionale è percepita come un'incombenza burocratica onerosa, e di fatto a volte lo è, soprattutto quando si devono compilare tante schede contemporaneamente, per cui non sempre vi si dedica sufficiente tempo. Occorre anche sottolineare che i repository istituzionali sono contenitori generici, in quanto comprendono le pubblicazioni di tutte le discipline presenti in un ateneo,

37 Kayvan Kousha; Mike Thelwall, *Artificial intelligence to support publishing and peer review: a summary and review*, «Learned publishing», 37 (2024), p. 4-12.

38 Con SSD - settore scientifico disciplinare - si intende una aggregazione disciplinare come definita dal Ministero dell'università e della ricerca. Per approfondimenti si veda <https://www.mur.gov.it/it/aree-tematiche/universita/docenti-universitari-e-carriera-accademica>.

39 67 atenei, cui si aggiungono 9 istituzioni, di ricerca secondo i dati Cineca del 2020.

40 Ministero dell'università e della Ricerca. *Decreto ministeriale n. 639 del 02-05-2024*. <https://www.mur.gov.it/it/atti-e-normativa/decreto-ministeriale-n-639-del-02-05-2024>.

e dunque, come dicevamo sopra, la parola 'biblioteca' o 'biblioteconomia' in un repository multidisciplinare può essere efficace.

L'attribuzione delle parole chiave poi dipende dall'approccio dell'autore, per cui alcuni considerano l'audience potenziale, altri cercano le *keyword* in articoli sullo stesso argomento di altri autori, alcuni hanno un approccio sistematico e studiato mentre altri inseriscono le prime parole che vengono loro in mente[41]. Le parole chiave attribuite a uno stesso articolo da ciascuno dei coautori in diversi repositories potrebbero quindi essere differenti tra loro, soprattutto se i coautori arrivano da settori disciplinari o da specializzazioni diverse.

La nostra analisi dei dati di IRIS ha il solo scopo di verificare quanto in profondità vadano le *keyword* per permetterci di trarre delle conclusioni e formulare delle proposte basate su dati effettivi. Siamo consapevoli, peraltro, dei limiti di un'estrazione dati dalle istanze IRIS degli atenei. IRIS presenta un campo specifico 'parole chiave' ma non in tutti gli atenei è un campo obbligatorio, di conseguenza non tutti gli autori devono attribuire delle *keyword* ai loro caricamenti nel repository. Negli atenei in cui è un campo obbligatorio, può essere richiesto un numero minimo di parole chiave oppure non essere indicato alcun numero minimo; dunque, una pubblicazione può avere una sola parola chiave oppure molte più di una o nessuna. Non sempre vi sono indicazioni prescrittive dell'ateneo riguardo alla lingua da preferire. Abbiamo quindi utilizzato il dato complessivo di tutte le lingue e le *keyword* sulle quali abbiamo lavorato possono essere in qualsiasi lingua. In assenza di linee guida specifiche, è lasciato all'autore decidere il numero di termini, per cui si può trattare di *keyword* oppure di *key phrases*[42]. I dati sono stati analizzati con lo strumento aperto di analisi e visualizzazione Voyant tools[43], tokenizzando le singole parole, con l'esclusione delle *stop words*, non significative. Non sono state usate le *key phrases*. Questo può essere un limite e vincola a un'analisi di un certo tipo, di cui bisogna essere consapevoli, perché si perdono i legami tra le parole, ma ci consente di enfatizzare i termini più usati e di rispettare maggiormente la restituzione in forma anonima dei risultati. La *word cloud* generata con Voyant Tools è rappresentata nella figura 2.

Figura 2 – Word cloud dei repository istituzionali italiani

41 A. Roy; S. Ghosh, *Freedom versus standard in article keyword generation* cit., p. 295.

42 Si veda la nota 7.

43 Voyant tools è un ambiente web per la lettura e l'analisi di testi digitali: <https://voyant-tools.org/>.

Come mostra la *word cloud*, le parole più frequenti nel corpus di 4893 record con il campo 'parola chiave' compilato (con una o più *keyword*s) sono: storia (1122); biblioteche (849); archivi (611). Le parole chiave che vengono aggiunte in IRIS possono corrispondere o meno a quelle inserite nella rivista in cui è pubblicato l'articolo, a seconda della scelta dell'autore o autrice.

La word cloud raccoglie le cinquanta parole più frequenti e mostra come si caratterizzino per l'ampiezza più che per la profondità. *Keyword* come 'biblioteconomia', 'bibliografia', coincidono totalmente o in gran parte con il campo semantico della disciplina. Come già accennato, i repository istituzionali sono contenitori generali che includono pubblicazioni in tutti i settori disciplinari presenti nell'ateneo, per cui in questo contesto una parola come 'biblioteconomia' diventa specifica, rispetto per esempio a un repository disciplinare o a una rivista del settore. Tuttavia, ai fini del recupero dell'informazione da parte di un utente specializzato, come per esempio uno studioso di scienze biblioteconomiche, oppure per un sistema di ASEO, diventa generica. Nella piramide di Lebrun le *keyword* generiche si collocano alla massima ampiezza e quindi alla minima profondità. Dalla nostra analisi risulta una certa disomogeneità nell'uso delle parole chiave da parte degli autori, cosa che può rendere difficoltoso il reperimento di articoli di interesse. Va anche detto che la consapevolezza della loro importanza è cresciuta nel tempo con l'evolvere delle tecnologie digitali, motivo per cui è probabile che alcuni risultati poco efficaci (*keyword* generiche o costruite come stringhe di soggetto) siano da riferire a inserimenti di pubblicazioni nei repository avvenute nei periodi iniziali del range temporale preso in esame. Occorre aggiungere che il bilanciamento ideale tra ampiezza e profondità non è affatto facile da ottenere a livello di singolo ricercatore. Al contrario è un obiettivo che dovrebbe essere perseguito con una strutturazione generale e un controllo centrale sulle parole chiave inserite dai singoli, magari svolto dalle biblioteche, come avviene negli esempi riportati nel paragrafo 5.

Proprio su questo genere di strutturazione fortemente centralizzate si basano invece le banche dati editoriali commerciali che raccolgono e indicizzano le riviste scientifiche. Tendenzialmente sono caratterizzate da contenuti editoriali in tante discipline diverse e risulta evidente il grande investimento in termini di organizzazione dell'informazione che gli editori e gli aggregatori portano avanti in queste banche dati e, specificamente per le *keyword*, per poter raggiungere un bilanciamento tra ampiezza e profondità che massimizzi il recupero dell'informazione da parte dell'utente. Abbiamo condotto alcune verifiche sulla banca dati citazionale Scopus (che è anche banca dati bibliografica) e sulla banca dati di business intelligence SciVal, entrambe prodotte da Elsevier, confrontando un campione di articoli tratti dalle riviste «AIB studi» e «JLIS.it». La verifica ha dimostrato l'approccio scrupoloso nell'attribuzione delle *keyword* in questa banca dati che reindicizza le *keyword* attribuite agli articoli dagli autori e presenti sulla rivista in modo da farle rientrare nella struttura generale di indicizzazione controllata centralmente. In particolare, Scopus mantiene le parole chiave dell'autore ma aggiunge le *keyword* che si riferiscono ai Topics di SciVal (si veda infra). Da notare come gli articoli di «AIB studi» non presentino le *keyword* dell'autore, che pure sono inserite negli articoli della rivista. Il motivo probabile è che le *keyword* sulla rivista online sono collocate in una pagina intermedia e non sono presenti sull'articolo in PDF (né nella versione HTML), a differenza di «JLIS.it» che invece include le *keyword* nel PDF. Il sistema di *harvesting* di Scopus evidentemente cattura solo i dati presenti sull'articolo che viene importato e non nelle pagine intermedie.

SciVal è una banca dati che usa i dati e i metadati di Scopus per fornire varie elaborazioni e visualizzazioni. Com'è noto, la banca dati citazionale di Elsevier è sbi-

lanciata verso i settori disciplinari STEM, per cui le *keyword* tendono a volte a far rientrare gli articoli nei macrosettori di interesse trasversale, ma non mancano *keyword* più specifiche come '*academic libraries*', '*public libraries*', '*librarian*'. Si rimane però a un livello intermedio, poiché termini quali 'controllo bibliografico' oppure '*data librarianship*' non sono presenti nelle *keyword* assegnate da Scopus. Inoltre, è interessante notare come le *subject areas* attribuite a «JLIS.it» siano tre, con pari quantità di articoli al loro interno, ovvero «Arts & humanities», «Computer science» e «Social sciences», mentre gli articoli di «AIB studi» rientrano solo nella subject area '*Social sciences*'. Eppure da una ricerca della *keyword* '*computer science*' nella piattaforma di «AIB studi» risultano ben 196 articoli. Ma '*computer science*' non rientra tra le *keywords* attribuite dagli autori e presenti negli articoli e forse questo non permette il loro recupero in SciVal.

Esempi di questo genere potrebbero essere tratti dai prodotti di vari altri editori scientifici. Ciascuna piattaforma commerciale organizza i contenuti in maniera mirata per poter offrire all'utente servizi avanzati che vanno dai sistemi di *alerting* ai suggerimenti di lettura di articoli simili a ciò che si sta leggendo, oggi sempre più basati sull'intelligenza artificiale. Le parole chiave associate agli articoli sono uno dei parametri utilizzati per alimentare gli algoritmi che stanno alla base di questi servizi ma funzionano solo se il sistema di attribuzione è organizzato e curato. Una funzionalità che potrebbe essere implementata anche nei repositories istituzionali, perfino a livello nazionale, visto che peraltro la gran parte degli atenei italiani utilizza lo stesso CRIS.

5. Il ruolo dei bibliotecari

Come evidenziato sopra, viste le innumerevoli funzioni delle parole chiave, gli editori, soprattutto i grandi editori commerciali, vi prestano crescente attenzione. Innanzitutto, forniscono indicazioni precise agli autori sui criteri di scelta delle parole chiave, come nell'esempio citato di Springer. Inoltre, nelle banche dati editoriali, organizzano i contenuti secondo sistemi altamente strutturati con delle *keyword* proprie basate su vocabolari controllati, come nel caso di Elsevier. La 'scheda' di un articolo recuperato nella banca dati Scopus presenta infatti sia le *keyword* inserite dall'autore nell'articolo sulla rivista (*author keywords*) sia le *keyword* assegnate da Scopus tramite un processo di indicizzazione standardizzato (*indexed keywords*) che, a differenza delle *keyword* dell'autore, prende in considerazione i sinonimi, le variazioni ortografiche, i plurali[44]. Le *keyword* sono uno degli elementi che contribuiscono a determinare la rilevanza di un articolo rispetto a una stringa di ricerca. Inoltre, in Scopus è presente una maschera di ricerca che consente di trovare i ricercatori che si occupano di tematiche affini tramite le *keyword*[45]. SciVal costruisce gli indici in maniera ancora più elaborata con algoritmi complessi nei quali, tuttavia, rientrano anche le *keyword*, sia per l'assegnazione dell'articolo ai *Topics* e *Topic clusters*, che sono collezioni di articoli con un focus intellettuale comune determinato tramite una serie di parametri in gran parte citazionali, sia per l'attribuzione di un SDG (Sustainable development goal) dell'Agenda 2030.

L'approfondimento di questi algoritmi ci porterebbero fuori tema ma è utile ancora sottolineare come Elsevier abbia creato il sistema di text mining Fingerprint Engi-

44 Scopus Support Center, *How do Author keywords and Indexed keywords work?* <https://service.elsevier.com/app/answers/detail/a_id/21730/supporthub/scopus/kw/*keyword*/>.

45 Researcher Discovery <https://www.scopus.com/search/form.uri?display=basic&zone=header&origin=searchbasic#researcher-discovery>.

ne che applica una serie di tecniche di Natural Language Processing (NPL) e che è stato allenato su dieci diversi thesauri disciplinari per poi diventare dal 2024 un thesaurus proprietario OmniScience[46]. Si tratta di un sistema dinamico che consente di superare alcuni dei limiti dell'indicizzazione dei contenuti editoriali, in particolare rispetto all'aggiornamento e soprattutto nelle discipline umanistiche dove vi sono grande varietà e specificità che evolvono nel corso del tempo.

I mezzi a disposizione dei grandi editori commerciali non sono ovviamente comparabili a quelli delle riviste gestite da piccoli editori, associazioni, università, che peraltro caratterizzano il mondo della ricerca umanistica più di quella STEM. Tuttavia, non dobbiamo dimenticare che una delle vie all'open access (la *green road to open access*) è basata sui repository istituzionali o disciplinari che proprio per tale ragione dovrebbero essere rafforzati con politiche mirate, come avviene in altri paesi.

Il caso della Svezia va proprio in questa direzione. La National Library of Sweden ha elaborato, su richiesta del governo, la *policy* per l'open access. Questa include linee guida per la gestione dei repository istituzionali degli atenei, e una lista controllata di *national subject categories,* alle quali l'autore che deposita una pubblicazione in un repository istituzionale deve fare riferimento per compilare un campo obbligatorio. I bibliotecari verificano le parole chiave inserite nel repository e introducono delle modifiche dove necessario. La modifica più frequente è la sostituzione della *keyword* dell'autore con una parola chiave più specifica[47]. Le *national subject categories* rimangono a livello disciplinare ampio e sono i termini indicizzati controllati della classificazione standardizzata svedese delle aree di ricerca. Esistono tre livelli di suddivisione gerarchica: sei categorie di livello superiore, suddivise in un totale di 42 categorie al secondo livello gerarchico, seguite da circa 250 categorie al livello più profondo. I primi due livelli corrispondono alla classificazione internazionale dei temi di ricerca dell'OCSE (Organizzazione per la cooperazione e lo sviluppo economico), mentre il terzo è specifico della Svezia. Lo scopo principale di queste categorie è facilitare la raccolta di dati per le statistiche nazionali ufficiali sulle pubblicazioni scientifiche. L'indicazione di almeno una *national subject category* nella scheda del repository è obbligatoria. Oltre alla *national subject category*, gli autori possono inserire nei repositories istituzionali altri due metadati: *university research subjects* e *keyword* libere. I primi sono i temi di ricerca secondo la dicitura e l'organizzazione presenti all'interno dell'ateneo dell'autore, che possono coincidere o non coincidere con le *national subject categories*[48]. È un campo opzionale che però se non compilato dall'autore viene integrato dai bibliotecari. Le *keyword* sono termini liberi che l'autore può inserire nel record. Trattandosi di un campo opzionale, se non viene compilato sono i bibliotecari a occuparsi di verificare se sono presenti delle *keyword* sul sito della rivista ed eventualmente copiarle. Se non sono presenti, il campo viene lasciato vuoto.

Un pionieristico progetto cooperativo analogo per i repository istituzionali era BPRESS, sviluppato nel 1999 come Berkeley electronic press, dal 2017 di proprietà del

46 Véronique Malaisé; Anke Otten; Pascal Coupet, *OmniScience and Extensions – Lessons Learned from Designing a Multi-domain, Multi-use Case Knowledge Representation System*, in *Knowledge Engineering and Knowledge Management. 21st International Conference, EKAW 2018, Nancy, France, November 12-16, 2018, Proceedings*, Springer, Cham, 2018, DOI: 10.1007/978-3-030-03667-6_15p. 228-242>.

47 K. Golub [*et al.*], *Subject indexing in humanities* cit.

48 Le differenze possono dipendere da motivi quali la presenza o meno di un settore di ricerca all'interno dell'ateneo, oppure dalla classificazione diversa a livello gerarchico dei settori; per i dettagli si rimanda all'articolo citato.

gruppo RELX (di cui fa parte Elsevier). La piattaforma *Digital commons* di BPRESS offriva una tassonomia dalla quale gli autori potevano selezionare le *keyword* che poi venivano controllate dal personale responsabile per il repository nei diversi dipartimenti[49].

Progetti che vadano in questa direzione diventano essenziali per il recupero dell'informazione nei repositories come per l'applicazione di strumenti di intelligenza artificiale, realizzabili auspicabilmente anche dalle stesse università e istituzioni pubbliche. Un'indagine di COAR (Confederation of open access repositories) pubblicata ad aprile 2025 mostra come su un campione di 66 repository internazionali, il 90% abbia registrato 'incursioni' di *bot* di intelligenza artificiale, anche più di una volta alla settimana[50].

Sono ormai molti i supporti alla ricerca basati sull'intelligenza artificiale. Per esempio, i sistemi di raccomandazione (*recommender systems*) che utilizzano l'IA per offrire suggerimenti di lettura in ambito accademico e scientifico filtrando i contenuti con algoritmi complessi basati su una combinazione di molti parametri, tra i quali le *keyword* rivestono un ruolo importante. Ne è un esempio il *recommender system* di CORE (COnnecting REpositories), organizzazione no-profit internazionale per l'implementazione dell'open access, che analizza ed estrae suggerimenti proprio dai repositories open access aderenti alla rete[51]. Un altro esempio sono le ormai numerose applicazioni che generano riassunti degli articoli, evidenziandone i temi principali, o quelle che scandagliano il web per creare delle rassegne bibliografiche su determinati argomenti[52], le cui architetture Large language models (LLM) possono avvantaggiarsi dalla presenza di *keyword* scelte con precisione. Le potenzialità che una maggiore organizzazione cooperativa e valorizzazione dei repositories istituzionali in Italia potrebbe offrire, sono indubbiamente significative.

6. Conclusioni

Risulta innegabile che le *keyword* hanno assunto un ruolo cruciale nella comunicazione scientifica e che la loro importanza, spesso sottovalutata sia dagli autori che dagli editori, sia fondamentale per la buona divulgazione dei risultati della ricerca e per la definizione degli ambiti e sotto ambiti disciplinari. Infatti, le *keyword* nell'editoria scientifica sono uno strumento fondamentale per garantire che la ricerca pubblicata sia facilmente trovata, compresa e citata dalla comunità scientifica e da altri potenziali lettori o utilizzatori, siano essi umani o artificiali.

I dati analizzati dimostrano che non sempre la scelta delle *keyword* è efficace e funzionale all'indicizzazione degli articoli scientifici e alla loro disseminazione. La

49 Maira Bundza, *The choice is yours! Researchers assign subject metadata to their own materials in institutional repositories*, «Cataloging & classification quarterly», 52 (2014), n. 1, p. 110-118, DOI: 10.1080/01639374.2013.8524.

50 COAR, *Open repositories are being profoundly impacted by AI bots and other crawlers: results of a COAR survey*, 30 April 2025, <https://coar-repositories.org/news-updates/open-repositories-are-being-profoundly-impacted-by-ai-bots-and-other-crawlers-results-of-a-coar-survey/>.

51 CORE recommender, <https://core.ac.uk/services/recommender>; si veda anche Petr Knoth; Zdenek Zdrahal, *CORE: connecting repositories in the open access domain*, <https://indico.cern.ch/event/103325/contributions/1300458/>.

52 Carlo Galli; Chiara Moretti; Elena Calciolari, *Intelligent summaries: will artificial intelligence mark the finale for biomedical literature reviews?* «Learned publishing», 38 (2025), n. 1, DOI: 10.1002/leap.1648; alcuni tra gli innumerevoli esempi sono Scholarcy <https://www.scholarcy.com/>, Humata <https://www.humata.ai/>, Scisummary <https://scisummary.com/>, e il tool Copilot integrato in Microsoft Office.

casistica riscontrata è decisamente varia, e può comprendere parole troppo generiche per essere veramente efficaci, uso indifferente di forme singolari o plurali, *keyword* fra loro collegate come stringhe di soggetto, cosa che fa sì che i singoli termini perdano il loro valore semantico, dato solo dal contesto della stringa stessa. A fronte di una situazione di questo genere, stranamente non esistono linee guida o proposte di buone pratiche, a parte quanto indicato nei siti degli editori commerciali, come già accennato precedentemente (cfr. paragrafo 2).

Come alcuni ricercatori hanno fatto per altri ambiti disciplinari, sarebbe forse opportuno proporre alla comunità scientifica di riferimento, pur nella sua ampiezza e nelle sue sfaccettature, un vocabolario controllato implementabile, in modo tale da facilitare la scelta delle *keyword* e garantire una certa uniformità nell'uso delle stesse. Un esempio in questo senso può essere quanto proposto già nel 1988 da Barki, Rivard e Talbot nell'articolo *An information systems keyword classification scheme*, aggiornato nel 1993[53], anche se si tratta di una proposta molto strutturata e che tende ad essere, ovviamente, quando dichiarato dal titolo, ovvero non un semplice thesaurus di termini ma un vero e proprio schema di classificazione, con tutte le caratteristiche del caso («A keyword classification scheme includes four components: schedules, a notation, an index, and an organization for maintaining and revising»[54]). Senza arrivare a questi livelli di approfondimento e di strutturazione, sarebbe però opportuno quantomeno sensibilizzare ricercatori e studiosi su queste questioni, che possono diventare problematiche nella misura in cui precludono l'accesso alle ricerche.

Per la proposta di buone pratiche per la realizzazione di *keyword*, è inevitabile utilizzare come punto di partenza quanto indicato dagli editori, gli unici a dare indicazioni facilmente reperibili su questo argomento. I suggerimenti dati dall'editore Springer sono già stati riportati nella parte iniziale dell'articolo[55]; nella loro semplicità possono essere un'ottima base per valutare l'efficacia delle *keyword* scelte. Una sintesi di quanto proposto da diverse riviste ed editori, sebbene si sia consapevoli del fatto che alcuni suggerimenti meglio si adattano a un determinato contesto e situazione che ad altri, può essere la seguente: le *keyword* per essere efficaci dovrebbero essere rilevanti, strettamente correlate al contenuto principale dell'articolo e specifiche, utilizzare la terminologia tecnica, includere la forma estesa nel caso si usino acronimi e sigle, utilizzare brevi frasi piuttosto che singole parole quando necessario; inoltre sarebbe opportuno analizzare articoli simili per prendere visione delle *keyword* scelte, e cercare di valutare quali potrebbero essere i termini utilizzati da coloro potenzialmente interessati all'articolo stesso. Ovviamente quanto detto sopra è molto essenziale; sarebbe invece opportuno un lavoro in sinergia fra produttori di *keyword*, editori, comunità scientifica di riferimento e bibliotecari al fine di individuare delle buone pratiche condivise su questi aspetti fondamentali, lo ribadiamo, della comunicazione scientifica.

In questo contesto, risulta fondamentale e necessario il ruolo attivo dei bibliotecari, accademici in primo luogo (già in parte affrontato in precedenza), e di colo-

53 Henri Barki; Suzanne Rivard; Jean Talbot, *An information systems keyword classification scheme,* «MIS quarterly», 12 (1988), n. 2, p. 299-322; Henri Barki; Suzanne Rivard; Jean Talbot, *A keyword classification scheme for is research literature: an update*, «MIS quarterly», 17 (1993), n. 2, p. 209-226.

54 H. Barki; S. Rivard; J. Talbot, *An information systems keyword classification scheme* cit., p. 300.

55 Per comodità riportiamo i punti essenziali: «Include your techniques and/or specific methodology; Focus on the main topic of your research; Avoid *keyword*s that are only one word; Avoid overlapping *keyword*s in your title and those in your *keyword* list; Follow the journal guidelines when selecting *keyword*s; Perform *keyword* research before submitting your article».

ro che si occupano di *knowledge organization*, innanzitutto perché dovrebbero avere le competenze per aiutare a migliorare le scelte fatte, e perché sono parte integrante della filiera del deposito dei prodotti nei repository istituzionali. Il ruolo dei bibliotecari potrebbe non limitarsi a questo, ma potrebbe spingersi oltre, sia a monte del processo, fornendo consulenza nella scelta delle *keyword*, nonché formazione e supporto ai ricercatori, sia migliorando la gestione delle *keyword* nei repository istituzionali, sia facendo da intermediario fra gli autori e i gestori di database, e infine fornendo una analisi e valutazione dell'efficacia delle *keyword* stesse.

Anche le associazioni di categoria e le società scientifiche di settore potrebbero essere parte attiva e integrante del processo perché una migliore indicizzazione aiuta tutti a far emergere i risultati della ricerca e a diffondere e disseminare quanto realizzato in ambito accademico (favorendo così anche la terza missione, o valorizzazione delle conoscenze, che è una delle funzioni fondamentali delle università).

Articolo proposto il 12 maggio 2025 e accettato il 7 luglio 2025.

ABSTRACT **AIB** studi, vol. 65 n. 1 (gennaio/aprile 2025), p. 00-00. DOI 10.2426/aibstudi-14149
ISSN: 2280-9112, E-ISSN: 2239-6152 - Copyright © 2025 Rossana Morriello, Lucia Sardo

ROSSANA MORRIELLO, Università degli studi di Firenze, e-mail: rossana.morriello@unifi.it
LUCIA SARDO, AMS Università di Bologna, lucia.sardo@unibo.it

L'uso delle *keyword* nella comunicazione scientifica, con un'analisi sul SSD M-STO/08 (HIST-04/C)
L'articolo esamina il ruolo delle keyword nella comunicazione scientifica, con un focus specifico sul settore disciplinare HIST-04/C (ex M-STO/08). Introdotte nell'editoria scientifica a partire dagli anni Settanta del Novecento, le keyword si sono affermate come strumenti essenziali per la ricercabilità, l'indicizzazione e la disseminazione delle pubblicazioni accademiche. L'analisi mette in luce vantaggi e criticità nella loro selezione, confrontandole anche con i sistemi di soggettazione in termini di capacità di richiamo e precisione, evidenziando al contempo l'assenza di linee guida condivise per la loro definizione. Si propone una più stretta collaborazione tra autori, editori, bibliotecari e comunità scientifica per l'elaborazione di buone pratiche e vocabolari controllati. L'articolo esplora inoltre il ruolo crescente dell'intelligenza artificiale e dei repository istituzionali, sottolineando la necessità di politiche adeguate e valorizzando il contributo dei bibliotecari nella gestione delle keyword. In conclusione, si evidenzia l'importanza delle parole chiave non solo per la diffusione della ricerca, ma anche per il coinvolgimento del pubblico non accademico e il sostegno alla terza missione dell'università.

Keyword Use in Scientific Communication: a focus on the disciplinary sector M-STO/08 (HIST-04/C)
This article examines the role of keywords in scientific communication, with a specific focus on the disciplinary field HIST-04/C (formerly M-STO/08). Introduced in academic publishing in the 1970s, keywords have become essential tools for enhancing the discoverability, indexing, and dissemination of scholarly publications. The analysis highlights both the advantages and challenges involved in keyword selection, also comparing them to subject indexing systems in terms of recall and precision, while pointing out the lack of shared guidelines for their formulation. The article advocates for closer collaboration among authors, publishers, librarians, and the academic community to develop best practices and controlled vocabularies. It also explores the growing role of artificial intelligence and institutional repositories, emphasizing the need for targeted policies and recognizing the librarians' contribution to keyword management. Finally, the article underscores the importance of keywords not only for research dissemination but also for engaging non-academic audiences and supporting the university's third mission.