

Zeitschrift: Schweizerisches Archiv für Volkskunde = Archives suisses des traditions populaires

Band: 68-69 (1972-1973)

Heft: 1-6: Festschrift für Robert Wildhaber zum 70. Geburtstag am 3. August 1972

Artikel: Per un soggettario demologico da costruire con impiego del computer

Autor: Cirese, Alberto M.

DOI: <https://doi.org/10.5169/seals-116772>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. [Siehe Rechtliche Hinweise.](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. [Voir Informations légales.](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. [See Legal notice.](#)

Download PDF: 17.11.2024

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

**Per un soggettario demologico da costruire
con impiego del computer**

di *Alberto M. Cirese*, Cagliari-Roma

Robert Wildhaber, cui qui rendiamo omaggio, ha dedicato una parte non trascurabile della sua lunga e preziosa attività di demologo ad una impresa bibliografica e documentaria che nel nostro campo è certo la più importante a livello internazionale: la *Volkskundliche Bibliographie*. Non gli riuscirà dunque sgradito, o almeno lo spero, se colgo proprio questa occasione per dare notizia del progetto di un repertorio di informazioni folkloriche ordinate per argomento o «soggetto» (e dunque di un «soggettario» come si usa dire più sbrigativamente), che di recente è stato avviato dalla cattedra di Storia delle tradizioni popolari dell'Università di Cagliari, in collegamento con il Centro Nazionale Universitario di Calcolo Elettronico (CNUCE) di Pisa, ed in parziale analogia con quanto si viene già facendo in altri settori d'indagine.

Come è doveroso in questa sede, ed anche in relazione al fatto che il lavoro è appena agli inizi, mi limiterò a fornire solo l'indicazione di alcune linee generalissime, riducendo al minimo i particolari tecnici e considerando inoltre la questione soltanto in relazione al materiale italiano. Confido comunque che quanto dirò basti per dare un'idea del progetto e per avviare eventualmente una collaborazione con chi già operasse nella stessa direzione.

* * *

Non ho bisogno di ricordare quali e quante siano le difficoltà che nel nostro campo si incontrano per reperire i dati documentari che restano dispersi in una vasta e sparpagliata messe di scritti. Le difficoltà sono poi tanto più gravi in Italia dove mancano totalmente opere di ricapitolazione complessiva quali sono invece disponibili per i paesi di lingua tedesca (*Handwörterbuch des deutschen Aberglaubens*; *Wörterbuch der deutschen Volkskunde* ecc.). Ne risulta che ogni volta che si abbia bisogno di rintracciare il materiale comparativo disponibile su di un qualsiasi argomento (p. es. l'accensione dei fuochi cerimoniali o falò, i modi con cui si contrae il legame del comparatico di San Giovanni, le forme dei pronostici ecc.), lo studioso è costretto a compiere una serie di spogli assai lunghi, gravosi e poco fruttuosi. Ci si deve applicare, ad esempio, a spogliare almeno gli indici delle venti e più annate dell'*Archivio per lo studio delle tradizioni popolari* di Pitre, e occorre poi ripetere l'operazione per le altre riviste fondamentali,

per gli Atti dei Congressi ecc. E ciò tante volte quanti sono gli argomenti studiati e quanti sono gli studiosi che di volta in volta si occupano dell'uno o dell'altro.

Sarebbe ovviamente più economico e insieme più fruttuoso se le operazioni di spoglio venissero condotte una volta per tutte: «a tappeto», come si suol dire, e cioè per tutte le pubblicazioni essenziali e per tutti gli argomenti emergenti. Ma la economicità *reale* di una simile impresa è strettamente legata alla quantità e alla qualità dei testi che si sottopongono allo spoglio, e inoltre dipende in modo diretto dal tipo di tecniche (manuali o automatiche) che si impiegheranno nello spoglio stesso.

Se ad esempio ci accingessimo a repertorializzare *tutte le informazioni* contenute in *tutte le pagine* anche delle sole riviste demologiche italiane, avremmo dinanzi un lavoro che supera i limiti della pazienza e della vita non di un solo ricercatore ma di intere *équipes*. Tanto più, poi, se intendessimo svolgere il lavoro *manualmente*.

Occorre quindi: *a)* ridurre la «lunghezza» complessiva dei testi da elaborare entro limiti che ne consentano la maneggevolezza e la esauribilità in tempi ragionevoli; *b)* sfruttare le possibilità di elaborazione *automatica* che sono oggi offerte dai calcolatori elettronici o computer.

Per cominciare da questo secondo punto, è ormai ben noto che i calcolatori elettronici sono perfettamente in grado di fornirci le «concordanze» di un qualsiasi testo: possono cioè darci automaticamente l'insieme dei «passi» (o «brani», «frasi», «contesti») che di volta in volta contengono la o le parole che ci interessano: per es. «*falò*», «*comparatico*», «*San Giovanni*», «*pronostico*» ecc. In linea di principio, dunque, nulla escluderebbe che per realizzare il repertorio delle informazioni contenute nell'*Archivio* di Pitre si passassero su schede perforate tutti gli scritti contenuti nell'*Archivio* stesso, e si richiedessero poi al computer le concordanze. Ma otterremmo così una mole spaventosa di dati che sarebbero per la massima parte irrilevanti ai nostri fini: anche escludendo automaticamente dalle operazioni le cosiddette «parole grammaticali» (congiunzioni, pronomi, articoli, preposizioni ecc.), resterebbe comunque una eccessiva massa di dati inutili ai fini della documentazione demologica. Infatti, anche in una rivista specializzata come quella di Pitre, la *densità* delle informazioni demologicamente rilevanti, per alta che sia, resta sempre *troppo bassa* per rendere realmente proficua una operazione a tappeto come quella ora indicata.

Si può essere allora tentati di procedere ad una preventiva scelta dei «passi» da sottoporre alla perforazione e alle concordanze, magari

anche operando la loro condensazione repertoriale, impiegando «parole-chiave» ecc. Ma la lunghezza complessiva del corpus continua a restare troppo grande perché queste scelte e queste condensazioni appaiano davvero realizzabili in tempi ragionevoli. Inoltre, dato che non è possibile prevedere tutte le eventualità che si presenteranno nel corso del lavoro, gli operatori sarebbero costretti ad una serie di decisioni che avverrebbero «caso per caso», e dunque anche in modo diverso per casi identici e in modo identico in casi diversi. Si perderebbe insomma gran parte dell'aiuto che le operazioni automatiche di concordanze possono fornire: altro è unificare le informazioni una volta che il calcolatore ci ha fornito *tutti i contesti* in cui una parola ricorre, ed altro è operare senza disporre di questa unificazione, sia pure grossolana e provvisoria.

Per sfruttare adeguatamente le possibilità offerte dalle elaborazioni elettroniche, dunque, occorre operare su testi che:

- a) siano i più *brevi* possibile;
- b) presentino il *massimo* di densità di informazioni;
- c) richiedano la quantità *minima* di interventi manuali-intellettuali nella fase di preparazione e di perforazione dei testi.

Gli «oggetti» che meglio sembrano corrispondere alle caratteristiche suddette sono i *titoli* degli scritti e, all'interno di ciascuno scritto, gli eventuali *sommari* o *indici* già forniti dagli autori.

Naturalmente ci sono molti titoli e molti indici che risultano generici e scarsamente indicativi; inoltre è evidente che anche il titolo o l'indice più preciso contiene solo una piccolissima parte delle informazioni complessive che sono viceversa contenute nell'articolo o nel libro. C'è insomma una forte perdita di informazioni, quando si operi solo sui titoli e sugli indici.

È questo però il prezzo che occorre pagare per rendere realizzabile l'impresa documentaria; ed un soggettoario, incompleto ma effettivamente realizzato, è sempre meglio di un soggettoario completo che non si può realizzare.

In ogni caso i titoli offrono una altissima densità media di informazioni. Se prendiamo ad esempio l'opera di Paolo Toschi, *Le origini del teatro italiano*, su 6 parole che compongono il titolo ben 3 sono portatrici di informazioni rilevanti, e cioè «*origini*», «*teatro*» e «*italiano*»; per giunta le tre restanti (e cioè «*le*», «*del*», e «*in*») sono «parole grammaticali» che possono essere eliminate automaticamente. E perciò le concordanze alfabeticamente ordinate risulterebbero come segue:

<i>Esponente</i>	<i>Contesto</i>
<i>italiano</i>	P. Toschi, Le origini del teatro ***
<i>origini</i>	P. Toschi, Le *** del teatro italiano
<i>teatro</i>	P. Toschi, Le origini del *** italiano

Posto che invece sottoponessimo a concordanze il titolo di un altro scritto dello stesso Toschi, *Sull'origine dello strambotto*, ne ricaveremmo:

<i>Esponente</i>	<i>Contesto</i>
<i>origine</i>	P. Toschi, Sull'*** dello strambotto
<i>strambotto</i>	P. Toschi, Sull'origine dello ***.

Su queste prime concordanze che diremo «grezze» si potranno poi ricavare agevolmente (e in parte anche automaticamente) le voci, estremamente ricche e articolate, del nostro soggettoario o repertorio tematico. Supponendo di indicare con T₁ e T₂ le due opere di Toschi addotte ad esempio, e supponendo anche di aver compiuto le operazioni di lemmatizzazione di cui si fa cenno più oltre, avremmo in ordine alfabetico:

<i>Italia</i>	teatro, origini T ₁
<i>Origini</i>	– dello strambotto T ₂ – del teatro italiano T ₁
<i>Strambotto</i>	origini T ₂
<i>Teatro</i>	origini, Italia T ₁ Italia, origini T ₁

Naturalmente le concordanze operate sul solo titolo dell'opera di Toschi dedicata al teatro non ci direbbero nulla sul fatto che nel libro si discorre a lungo del Carnevale. Ma se supponiamo di includere nel nostro lavoro di elaborazione anche il ricco *indice* dell'opera, siamo in grado di mettere in luce una massa di informazioni che certo è inferiore alle informazioni contenute nel volume ma è anche enormemente superiore a quelle fornite dal solo titolo. Ed anche gli indici, come si sa, sono testi abbastanza brevi e forniti comunque di una altissima densità di informazioni.

Ma per giungere dalle concordanze grezze al soggettoario vero e proprio occorre passare a quel lavoro manuale-intellettuale (e solo in parte automatico) che genericamente si chiama «lemmatizzazione». Nel nostro caso, come è naturale, si tratterebbe non tanto di una lemmatizzazione linguistica quanto invece di una lemmatizzazione *tematica*: per esempio le forme «Italia» e «italiano» andrebbero unificate, così come si dovrebbero unificare le forme «teatro» e «azione

drammatica» o simili. Ma l'intervento intellettuale-manuale (diretto a riunire ciò che è tematicamente identico o affine al di là della differenza linguistica, ed a dividere ciò che è tematicamente diverso al di là della identità di forma linguistica) ora avverrebbe su un materiale già larghissimamente sgrossato e reso enormemente più maneggevole.

Per facilitare l'operazione di lemmatizzazione tematica si possono operare alcuni agevoli e limitati interventi manuali sui testi, prima di passarli alla perforazione. Sembra infatti opportuno stabilire una ristrettissima serie di «informazioni privilegiate», codificandole preventivamente, in modo che il computer sia in grado di riconoscerle e quindi di selezionarle a parte. Sembra anche opportuno stabilire preliminarmente alcune categorie di informazioni che viceversa sono sicuramente irrilevanti, e che dunque intendiamo escludere preventivamente dalle nostre elaborazioni documentarie (pur non perdendo, però, la possibilità di recuperarle quando occorra).

Nel nostro lavoro, ad esempio, abbiamo convenuto di considerare «irrilevanti» le indicazioni bibliografiche esterne: luoghi di stampa, anno di edizione ecc. Basterà che queste indicazioni vengano perforate su schede di un certo «tipo» perché il calcolatore sia in grado di darci o di *non* darci le relative concordanze, a seconda dei nostri desideri o delle nostre necessità.

Per quanto riguarda invece le informazioni che abbiamo chiamato «privilegiate», si tratta essenzialmente delle seguenti (i termini impiegati per indicarle vanno presi in senso lato e convenzionale):

1) *persone* o *personaggi*, con distinzione tra *persone-agenti* (autori, curatori, prefatori di opere ecc.) e *persone-argomenti* (ossia individui reali o immaginari che costituiscono l'*oggetto* dell'opera);

2) *localizzazioni*, costituite sia da nomi di località, sia da «aggettivi» geografici, etnici ecc.

3) *datazioni*, siano esse quelle indicate con date più o meno precise, siano invece quelle più generiche del tipo «antico», «medievale», «moderno» ecc.

Si potranno così ottenere dal calcolatore delle concordanze circoscritte a ciascuna delle categorie sopra indicate, con evidente facilitazione delle operazioni di lemmatizzazione tematica.

In verità nel nostro lavoro, oltre a precisare più esattamente le caratteristiche e i confini delle categorie privilegiate che abbiamo indicato, abbiamo preso in considerazione anche altri due tipi di informazioni da codificare preventivamente, e li abbiamo convenzionalmente denominati «*brani*» e «*titoli*», aggiungendo al tutto anche

una categoria di «*varie*» per i casi impreveduti o equivoci. Ma qui non ne parlerò per non appesantire ulteriormente una esposizione che lo è già troppo.

Concluderò invece richiedendo al lettore uno sforzo di immaginazione: quale sarebbe il frutto di un simile progetto se, superate le non insormontabili difficoltà relative alla presenza di titoli in molte lingue diverse, si riuscisse a realizzare l'impresa su un'opera così importante come la *Volkskundliche Bibliographie* alla quale tanto direttamente si lega l'operosa vita di Robert Wildhaber?