

La metrologia nella società della comunicazione: qualche riflessione

Luca Mari

Tutto_Misure, 1, 2026

[26.1.26]

Viviamo nella società della comunicazione, e una parte rilevante delle enormi quantità di dati che stiamo producendo – ed elaborando, trasmettendo e memorizzando – è il risultato di misurazioni, spesso realizzate in modo automatico o semi-automatico a partire dai sensori diffusi sempre più ovunque, dalle linee di produzione industriale ai dispositivi indossati dalle persone. Ha perciò senso chiedersi se la metrologia – “scienza della misurazione e delle sue applicazioni”, in accordo al Vocabolario Internazionale di Metrologia (VIM) – sia in qualche modo influenzata, e nel caso come, dall’importanza che la comunicazione ha progressivamente assunto. Quanto segue è un’introduzione a qualche aspetto di questo tema.

Big data o big information?

A proposito di *big data*, in un’intervista del 2010 Eric Schmidt, allora amministratore delegato di Google, dichiarò: “Il mondo non è pronto per la rivoluzione tecnologica che presto accadrà. Sia perché le persone non capiscono quello che sta per succedere, sia per le cose stesse che stanno succedendo: sono stati prodotti cinque exabyte di informazione dall’inizio della civilizzazione al 2003, e ora la stessa quantità viene prodotta ogni due giorni. Non c’è da stupirsi che possiamo sentirci schiacciati da questo.” (<https://m.youtube.com/watch?v=UAcCIsrAq70&t=467s>).

Che nel 2010 fossero proprio cinque exabyte, e che oggi non siano più cinque, ma dieci, cinquanta o cinquecento, poco importa. Importa invece che in un periodo assai breve questa quantità è aumentata di molti ordini di grandezza (almeno un milione di volte, facendo qualche conto approssimativo sui numeri proposti da Schmidt), un cambiamento quantitativo che inevitabilmente diventa anche qualitativo: quello che sta accadendo è non solo *un di più* rispetto a un passato anche recente, ma anche e soprattutto *un qualcos’altro*. E importa chiarire il genere di entità a cui queste stime fanno riferimento: si tratta davvero di una certa, enorme, quantità di informazione? E perché allora il termine “big data”, e non “big information”?

(Una nota linguistica. In inglese “information” è non numerabile: si dice per esempio “some information”, non “informations”. In italiano “informazioni” è corretto e comune, ma quando il significato è tecnico si tende a replicare l’inglese e a usare “informazione” in forma non numerabile: per esempio, “si misura per produrre informazione sul misurando”, e non “informazioni”. L’inglese sta progressivamente trattando “data” come un sostantivo non numerabile, e perciò “data is...”, o comunque dimenticando il singolare “datum” di derivazione latina, e perciò se anche si mantiene “data are...” si evita “datum is...”. In italiano stiamo invece mantenendo la distinzione tra “dato” e “dati”).

Per comprendere il senso stesso della misurazione *la differenza tra dati e informazione è cruciale*.

Ripartendo da Shannon

Come di punto di partenza per un’analisi a questo riguardo prendiamo il testo intitolato *Recent contributions to the mathematical theory of communication*, che Warren Weaver scrisse in apertura a *The mathematical theory of communication*, il libro pubblicato nel 1949 che contiene l’omonimo articolo di Claude Shannon, fondativo per quella che è stata chiamata “teoria dell’informazione”. In quel testo Weaver distingue tre “livelli” di problemi della comunicazione:

“Livello A: quanto accuratamente possono essere trasmessi i simboli della comunicazione?” (il problema tecnico)

Livello B: quanto precisamente i simboli trasmessi portano il significato inteso? (il problema semantico)

Livello C: quanto efficacemente il significato ricevuto influenza il comportamento nel modo desiderato? (il problema dell’efficacia)”.

Nei suoi commenti, Weaver riconosce esplicitamente che la soluzione almeno parziale dei problemi di un livello è una condizione per affrontare i problemi dei livelli successivi, e anche con ciò giustifica l'importanza di lavorare sul Livello A, quello che in effetti la teoria di Shannon fa. E poi introduce una spiegazione determinante per il nostro tema: "Il termine informazione, in questa teoria, è impiegato in un senso speciale e non è perciò da confondere con il suo uso abituale. In particolare, l'informazione non deve essere confusa con il significato. Infatti, da questo punto di vista, due messaggi, uno ricco di significato e l'altro completamente senza senso, possono essere esattamente equivalenti quanto all'informazione che portano." Su questa base si comprende il lessico della teoria di Shannon, per cui per esempio un'entità che emette simboli verso un canale di trasmissione è chiamata "sorgente di informazione" e l'entropia è la quantità media di informazione portata dai simboli emessi.

In accordo alla tripartizione per livelli di Weaver, se ne può concludere che il termine "informazione" è usato con tre accezioni diverse: (A) informazione senza significato, (B) informazione con significato, (C) informazione con significato ed efficacia. A quale accezione faceva allora riferimento Schmidt parlando di cinque exabyte di informazione prodotti in due giorni? Non ci dovrebbero essere dubbi, ma la questione diventa ancora più chiara se interpretiamo le condizioni che distinguono i "livelli" di Weaver – presenza di significato nel messaggio e sua efficacia rispetto a un certo scopo – alla luce della semiotica.

(A) Informazione senza significato, e dunque *informazione sintattica*.

A questo livello una sorgente è caratterizzata da:

- l'insieme dei simboli che può emettere, cioè l'alfabeto della sorgente – termine usato da Shannon stesso – che per semplicità supponiamo discreto e finito; ciò che è rilevante è la possibilità di scelta tra almeno due simboli diversi: se infatti l'alfabeto contenesse un solo simbolo la sorgente non potrebbe che emettere quel simbolo, un evento che non porta informazione;
- l'eventuale probabilità di emissione di ogni simbolo (di nuovo secondo Weaver, "l'informazione nella teoria della comunicazione ha a che vedere non tanto con quello che si trasmette, quanto con quello che potrebbe trasmettere: l'informazione è una misura [è scritto proprio così: "information is a measure"] della libertà di scelta nella selezione di un messaggio");
- le regole per la composizione delle successioni di simboli, il cui rispetto è preconditione perché una successione possa avere un significato (per esempio la successione di simboli "1,2,3,4*/+" non significa nulla perché non è sintatticamente "ben formata", pur essendo costituita da simboli dell'alfabeto della matematica).

(B) Informazione con significato, e dunque *informazione semantica*.

Cosa sia il significato è una questione non ovvia, e al proposito è stato scritto tanto (tra cui ovviamente degli articoli intitolati "The meaning of meaning"...). Anche per evitare di entrare in discussioni filosofiche, ci è sufficiente considerare una semantica come una relazione tra due sintassi, una relativa all'alfabeto dei simboli che una sorgente può emettere e una relativa ad altre entità, che stabiliscono appunto il significato dei simboli (per esempio la successione di simboli matematici ben formata "100" designa abitualmente il numero cento, ma può anche significare per esempio il numero quattro, se interpretata come costituita di simboli binari).

In questo contesto, la relazione è spesso chiamata un "codice", e la conoscenza del codice della sorgente è perciò condizione perché il destinatario sia in grado di decodificare il messaggio – in pratica calcolando l'inverso della funzione di codifica – e ricostruirne così il significato.

Si noti che l'informazione semantica non coincide con la verità del contenuto dei messaggi, perché un messaggio può essere significativo anche se falso (per esempio "1+1=10" è falso se interpretato in codice decimale, ma sappiamo che è falso perché lo comprendiamo) e perché ci sono messaggi significativi che non sono né veri né falsi (per esempio le definizioni).

(C) Informazione con significato ed efficacia, e dunque *informazione pragmatica*.

Una successione di simboli, che sia ben formata rispetto a una certa sintassi e dotata di significato rispetto a un certo codice, può essere più o meno efficace per lo scopo per cui è stata prodotta. Questo terzo livello coglie la componente pragmatica dell'informazione, che ha a che vedere con l'utilità dell'informazione stessa: mentre un messaggio a cui non si sa attribuire un significato non è né utile né inutile, un messaggio significativo è caratterizzato da quanto è utile, e quindi in particolare da quanto le risorse impiegate per produrlo sono giustificate.

Un'ipotesi

Questa analisi ci conduce a un'ipotesi: pur senza la pretesa di esaurire la ricchezza e la complessità del significato di questi termini, *“dato” è interpretabile come sinonimo di “informazione sintattica”, e “informazione” è interpretabile come sinonimo di “informazione semantica”* (in riferimento alla cosiddetta “piramide dati-informazione-conoscenza”, si potrebbe completare l'ipotesi sostenendo che in questo contesto “conoscenza” è interpretabile come sinonimo di “informazione pragmatica”: ma quest'ultimo passaggio ci porterebbe troppo lontano, e quindi non lo sviluppiamo qui).

Proviamo a esplorare qualche aspetto di questa ipotesi.

Primo, questa ipotesi rende conto della differenza tra fatti e dati: a differenza dei fatti, i dati non “sono nel mondo”, essendo condizionati da una decisione da parte di un soggetto. Che in un certo momento una certa lampadina sia accesa è un fatto, ma diventa un dato solo perché si è riconosciuto che la lampadina avrebbe anche potuto essere spenta, e in conseguenza si è deciso che l'alfabeto di questa comunicazione sia costituito dai simboli “acceso” e “spento” (o “0” e “1”, ecc.).

Secondo, questa ipotesi rende conto dell'idea che ci possono essere dati che non portano informazione, e che invece per esserci informazione ci devono essere dati che portano tale informazione. Che in un certo momento si riceva il simbolo “0” invece del simbolo “1” è un dato, che porta 1 bit di informazione sintattica se la probabilità attribuita a priori all'evento è $1/2$, ma non è ancora informazione. Lo diventa quando per esempio conosciamo il codice per cui “0” significa “quella certa lampadina in questo momento è spenta”.

Terzo, la teoria della comunicazione di Shannon è dunque una teoria della quantità di “datità” (di informazione sintattica, appunto) che una sorgente può emettere e un canale può trasmettere. Non è invece una teoria dell'informazione (nel senso di: informazione semantica), perché tratta di un sistema sintattico e non della relazione tra due sistemi sintattici, uno di significanti e uno di significati.

La conseguenza è ovvia: quello che caratterizza la nostra società è prima di tutto *un fenomeno di big data*, e non necessariamente di *big information*.

E la misurazione? Come contribuisce dunque a tutto ciò? E lo fa producendo dati o anche informazione?

Un tentativo di risposta a queste domande sarà il prossimo passo della nostra analisi.