

Classificazioni, ipotesi e software per il diritto

Paolo Bottazzini

Inquadramento del problema e logica dell'ipotesi

Il senso comune dipinge l'attività degli specialisti del diritto, della medicina, della finanza, come una missione che può essere intrapresa soltanto da individui che siano dotati per natura di un "fiuto" particolare, di un "occhio clinico" che permette loro di risolvere il rompicapo posto da ciascuna circostanza con l'intuizione che distingue il talento dal dilettante. Questo *istinto di verità*¹ che guida taluni a fare luce nell'intrico dei sintomi della situazione particolare elaborando al momento opportuno l'ipotesi corretta – è ciò che ha reso spesso i professionisti di questi settori eroi celebrati del cinema e della televisione. Negli ultimi anni tuttavia queste classi di specialisti hanno affidato all'interazione con motori di ricerca le fasi delicate dell'analisi dei sintomi e dell'elaborazione dell'ipotesi.

Consideriamo in particolare il settore della Giurisprudenza, restringendo il focus dell'attenzione sul segmento degli avvocati che si occupano di diritto civile. Le stime degli editori specializzati assegnano a quest'area professionale circa il 60-70% dei 180.000 avvocati iscritti all'albo in Italia. La chiave di volta del loro valore professionale si concentra nelle fasi che si succedono al momento delicatissimo dell'apertura di una nuova pratica. Il punto di innesco della storia che occorre analizzare è rappresentato dall'istante in cui il cliente presenta all'avvocato il problema sul quale invoca il suo patrocinio. Formalizziamo il processo in 4 fasi:

1. Il professionista riceve dal cliente l'insieme dei *dati* che devono permettergli di ricostruire l'*identità* del *caso giuridico*, tramite un'esposizione dei *sintomi* del problema lacunosa (dal punto di vista del diritto e spesso del buon senso), ed espressa in linguaggio quotidiano (ovvero non tecnico, sempre dal punto di vista del diritto);
2. i sintomi del problema autorizzano una serie di *ipotesi* relative al *principio di diritto* che lo disciplina;
3. l'ipotesi corretta permette l'identificazione del *testo normativo* (articoli dei Codici o delle Leggi Speciali) che esprime il principio di diritto esatto; la consultazione di Commenti e Rassegne consente di verificare in che modo la *tradizione interpretativa della giurisprudenza* abbia segmentato la formulazione astratta del principio di diritto in una *classificazione* di applicazioni su *casi concreti*;
4. ora l'avvocato dispone di una identificazione del caso giuridico in grado di motivare la compresenza dei sintomi elencati dall'esposizione del cliente, nonché di autorizzarne una conversione coerente, completa e corretta, in una esposizione linguistica giuridica. A questo punto l'avvocato è in grado di avviare la strategia e la procedura più opportuna per patrocinare il cliente.

¹ L'espressione è di C.S. Peirce: cfr. *Collected Papers*, 1931-1958 The Belknap of Harvard University Press, 7.219 (tr. it. a cura di M. Bonfantini, R. Grazia, G. Proni, *Storia e abduzione*, in *Le leggi dell'ipotesi*, 1984, Bompiani, pag. 274).

Il tipo di inferenza logica che viene applicata nel corso dell'interpretazione coincide con quella definita *abduzione*² da C.S. Peirce. Il percorso muove dalla verifica di un risultato, rappresentato dai dati empirici che vengono esposti dal cliente; passa attraverso la regola, identificata dal principio di diritto; termina con la scoperta della corretta identificazione del caso nella classificazione della giurisprudenza.

Gli eventi che si trovano in relazione con il risultato logico sono infiniti: sebbene la maggior parte di essi possano sembrare bizzarri al nostro gusto per la serietà della ricerca scientifica, non esistono ragioni logiche per escluderli da un'indagine esplicativa; e un computer non *esplicitamente* istruito³ non saprebbe farlo. La congiunzione tra risultato e regola si delinea come l'effetto di una *valutazione di pertinenza* tra le possibili spiegazioni della situazione – si profila come una *decisione*, che si trova iscritta nell'ambito di quanto appare *ragionevole* per la comunità umana o scientifica, ma che non è derivabile da alcuna forma di certezza meccanica. La valutazione inquadra la complessità del problema e “pesa” i singoli elementi rispetto alla comprensione del tutto; quindi seleziona il fattore più *rilevante*, esiliando gli altri temi sullo sfondo e sollevandolo al rango di regola. Si tratta del processo che Kant ha chiamato Giudizio riflettente, e che ha descritto come una facoltà non insegnabile – ma spontaneamente presente o assente in ogni individuo⁴.

La regola che deve essere individuata dall'avvocato che prende in carico il problema posto da un cliente è il principio di diritto che governa il caso concreto. Oltre alle difficoltà imposte dalla struttura dell'inferenza logica, il professionista giuridico deve confrontarsi con la selva legale italiana e le complessità che essa oppone alla sua ricerca di una soluzione. La tradizione interpretativa dispiegata dalla giurisprudenza non permette di fissare in modo univoco i canoni di applicazione dei principi di diritto – aprendo di fatto un ambito di indeterminatezza tra modalità di giudizio concorrenti. L'esito è che ciascun Tribunale e la Cassazione inaugurano ciascuno un proprio *orientamento* giurisprudenziale, organizzato in una versione maggioritaria e in proposte minoritarie. Sono questi impedimenti a motivare l'intervento provvidenziale del software e i requisiti della sua architettura.

Il ruolo delle piattaforma software

Nessuno sa con precisione quante leggi siano in vigore in Italia: alcuni stimano un numero complessivo di 300.000, altri 150.000, altri ancora 50.000. In ogni caso si

² Cfr. in particolare C.S. Peirce, *Deduction, Induction and Hypothesis*, in *Popular Science Monthly*, vol. 13, 1878 (tr. it. cit., *Deduzione, Induzione e Ipotesi*, in *Le leggi dell'ipotesi*, cit., pag. 205-220).

³ Sul tema della programmazione esplicita nel suo rapporto con il tema della rappresentazione della conoscenza, e la capacità di trarre inferenze dalle “conoscenze di Sfondo” o dai frame cognitivi, cfr. soprattutto la discussione esaminata in D.H. Hofstadter, *Gödel, Escher, Bach: an Eternal Golden Braid*, 1979, Basic Books, Inc (tr. it. a cura di Giuseppe Trautteur, *Gödel, Escher, Bach: un'Eterna Ghirlanda Brillante*, 1984 Adelphi; in particolare: cap. X, pag. 332-335; cap. XVIII, pag.665-676).

⁴ I. Kant *Erste Einleitung in die Kritik der Urteilskraft*, a cura di N. Hinske, W. Müller-Lauter, M. Theunissen, Stuttgart-Bad Cannstatt, 1965; (tr. it. a cura di P. Manganaro, *Prima introduzione alla Critica del Giudizio*, Laterza, Bari 1969).

tratta di una somma molto superiore a quella calcolabile negli altri Paesi europei, qualora si consideri che in Francia e in Germania i testi normativi vigenti sono meno di 10.000⁵. Occorre inoltre ricordare che la Corte di Cassazione Civile deposita ogni anno circa 30.000 sentenze⁶.

Tutti gli Studi di avvocatura in Italia hanno sottoscritto l'abbonamento ad almeno una piattaforma digitale per la consultazione delle sentenze depositate dalla Cassazione e di parte della giurisprudenza di merito. L'indagine che viene qui riassunta è stata condotta su circa 60 avvocati: l'intervistatore simula il comportamento di un potenziale cliente che dichiara di star ristrutturando il proprio appartamento e di voler aprire una finestra che si affaccerà su una strada privata; questa è larga 1 metro e viene regolarmente usata per il passaggio, in quanto congiunge due strade pubbliche (l'usanza è confermata anche dal piano urbanistico del comune). Il proprietario però è il vicino di fronte; il cliente domanda se può aprire la finestra senza chiedere la sua autorizzazione. Il problema si risolve consultando la successione di articoli del Codice Civile dal 900 al 907: sono le norme con le quali si apre la Sezione VII, Capo II, Titolo II del Libro Terzo, dedicato alle *luci e vedute*. L'Art. 900 si premura di spiegare che le *luci* e le *vedute* traducono nel dizionario tecnico del diritto ciò che nel linguaggio quotidiano viene genericamente denotato con *finestra*.

Gli avvocati intervistati non conducono però un'esplorazione preliminare del testo normativo, per verificare successivamente le interpretazioni dei precedenti giurisprudenziali; al contrario, essi si rivolgono immediatamente al motore di ricerca della piattaforma software imputando query che riprendono i lemmi del problema così come questo è stato formulato dal cliente. Oltre il 90% del campione interroga il motore dell'Archivio con stringhe del tipo `|finestra&strada&proprietà&vicino|`: gli avvocati selezionano i termini rilevanti del discorso del cliente traendoli dal dizionario del linguaggio comune e contando sulla sua sostanziale affinità con quello tecnico; al software viene consegnato l'incarico di rintracciare un insieme di risposte che permettano di derivare la traduzione tra i due vocabolari e di enucleare i criteri di classificazione appropriati.

Gli oggetti che il professionista consulta tramite il motore di ricerca sono le *mas-sime* estratte dalle sentenze: il loro formato editoriale è ridotto a poche righe di testo, in cui viene esposta una sintesi del caso concreto e si enuncia il principio di diritto invocato dall'autorità giudicante per emettere il verdetto. L'avvocato quindi cerca sentenze su casi simili a quello che gli è stato sottoposto, assumendo di poter risalire per analogia alla classificazione opportuna del caso e alla sussunzione sotto il principio corretto – purché la massima sia abbastanza recente. La tutela del *precedente* si accompagna per di più all'autorità dell'applicazione di un determinato principio di diritto da parte della Cassazione: una condizione che sembra mettere al riparo da errori di valutazione soggettiva e da contestazioni in aula.

Le dichiarazioni di entusiasmo per l'avvento dell'informatica in campo giuridico che vengono rilasciate dai soggetti intervistati al termine dell'operazione di test sono giustificate dal confronto con lo strumento tradizionale a loro disposizione: i Codici

⁵ La valutazione risale al 1999: se ne trova traccia in un articolo di Francesco Brugaletta, tuttora rintracciabile alla URL <http://www.apogonline.com/webzine/1999/12/13/01/199912130101>.

⁶ La stima è derivata dai dati statistici pubblicati sul sito web della Corte di Cassazione: cfr. <http://www.cortedicassazione.it/DocumentiPrimaPag/statistiche/Civile/StatisticheCivile.asp#>

Commentati, le Rassegne di Giurisprudenza e i fascicoli delle riviste – tutto in formato cartaceo. La ricerca delle sentenze con questi strumenti poteva comportare il dispendio di giorni, o settimane; era coordinata dagli Indici analitici della Giurisprudenza, il più prestigioso dei quali, redatto dal prof. Renato Borruso, prevede 4 gradi gerarchici di classificazione e 1.700 voci complessive al grado più basso. Ad un sistema così intricato di informazioni era possibile sopravvivere solo elaborando preventivamente una strategia di ricerca; ma anche questa poteva non bastare, dal momento che la struttura gerarchica ad albero richiede l'improbabile insorgenza nell'utente di un'empatia con l'autore della tassonomia che assicuri l'armonia prestabilita nell'assegnazione e nell'interpretazione delle etichette. Tuttavia, questa barriera di difficoltà permetteva al procedimento tradizionale di forzare nel professionista la generazione di una cultura completa intorno all'argomento esaminato – che implicava anche l'insieme delle scoperte casuali emergenti dalla serendipità dello svolgimento della ricerca.

Nella valutazione delle prestazioni del motore di ricerca, la riduzione del parametro di *precision* in favore del criterio di *recall*, nonché l'assenza di un ordinamento di *ranking*, assegna alla proprietà di linguaggio e all'accuratezza descrittiva dell'utente l'incarico di ridurre al minimo il *rumore informativo* che inquina l'elenco delle risposte. La completezza e la congruenza dei risultati selezionati dall'archivio delle massime rispetto alle esigenze conoscitive del professionista è quindi direttamente proporzionale alla correttezza dei lemmi iscritti nella stringa di ricerca secondo le regole dettate dal dizionario tecnico del diritto, e alla loro esaustività nell'espressione degli elementi salienti del problema. In mancanza di una interrogazione preliminare del testo normativo – unico vero serbatoio e depositario della tassonomia tecnica del diritto – la corrispondenza auspicata tende a ridursi al minimo. Nell'esempio proposto durante l'intervista agli avvocati la soluzione del quesito sarebbe emersa senza difficoltà sottoponendo al motore di ricerca una stringa di ricerca che avesse tradotto il lemma quotidiano *|finestra|* con l'espressione giuridica *|luce|* o *|veduta|*. La consuetudine di trascurare questo passaggio e l'inefficienza del software sono la causa della percentuale elevatissima di fallimenti registrata dal test.

Tre soluzioni e una riflessione

La giurisprudenza si è mostrata come la scienza tassonomica dei comportamenti sociali. Il reticolo concettuale della classificazione si offre ai professionisti come un sistema di ordinamenti e di procedure perfettamente compiuto; ciò di cui essi necessitano è una regola che permetta loro di sussumere gli eventi concreti ad una delle categorie disponibili. Si è visto in precedenza che il procedimento per rintracciarla non si delinea come un atto di abduzione diretto, ma passa attraverso il confronto con casi simili. Il comportamento dell'avvocato somiglia a quello del medico, che deve discernere l'identità della malattia tra le descrizioni imprecise del paziente, l'ambiguità dei sintomi, l'intersezione di patologie sincroniche, la storia sanitaria personale del soggetto. La collezione dei casi clinici incontrati durante la carriera gli permette di individuare gli elementi invarianti con cui confortare l'anamnesi del caso presente. Anche il procedimento degli avvocati è volto a classificare fatti, avvenimen-

ti – non cose. L'evento viene presentato da una sequenza disordinata di sintomi, che raffigurano circostanze, contesti sociali, frammenti di azioni, intenzioni. Questi dati devono essere raccolti come *esempi* di una classe di fatti, che sono identificati in quanto *casi di applicazione* di un *principio di diritto*. Il software compie tale lavoro di classificazione attraverso un esame delle occorrenze lessicali nella stringa imputata. Dalla varietà di software attualmente disponibili in commercio, o ad uso esclusivo della magistratura, è possibile enucleare un modello sincretico (d'ora in poi *MS*), la cui efficienza è ottimizzata in relazione alle esigenze informative dei professionisti esposte in precedenza. L'architettura di *MS* presenta l'insieme delle risposte alla query dell'utente con un'articolazione in 3 campi.

In primo luogo esegue un'*analisi spettrale* della collezione di item che corrispondono alla richiesta del professionista: questo processo permette di *classificare* le risposte sulla base dell'articolo del testo normativo citato nel corpo della massima. Il listato espone quindi in prima istanza l'elenco degli articoli del codice e delle Leggi Speciali invocate dalle autorità giudicanti nelle sentenze rintracciate; le sentenze appaiono come un attributo dei principi di diritto così identificati. Nel nostro esempio relativo alle finestre, il listato proporrebbe gli articoli 900 e 907 del Codice Civile associati al gruppo più numeroso di sentenze utili; citerebbe inoltre l'articolo 1120, relativo al decoro architettonico degli edifici, con un seguito più contenuto di sentenze; ecc. Il numero delle sentenze inviterebbe l'avvocato a considerare gli articoli 900 e 907 come particolarmente rilevanti per il tema che sta esplorando; ma il titolo dei testi normativi eserciterebbe una guida ancora più efficace nella selezione degli ambiti pertinenti per la ricerca, dal momento che l'articolo 900 disciplina la classificazione delle finestre, mentre il 907 governa il tema delle distanze dalla proprietà del vicino. Questa catalogazione *on-the-fly* è quindi in grado di offrire due chiavi euristiche: i testi normativi cui sono attribuite un numero maggiore di massime individuano i nuclei principali del problema; i titoli dei testi normativi permettono all'utente di dirigersi verso i temi più affini alla circostanza esposta dal cliente, escludendo già a priori tutte le sentenze collegate ad argomenti non pertinenti. Il delta che si era spalancato tra l'esigenza di massimizzare il parametro del *recall* a detrimento della *precision* viene ridotto al minimo con questo espediente.

In secondo luogo, l'architettura di *MS* esamina le *selezioni contestuali* della query e del dizionario che compone il testo dell'articolo normativo. Ogni lemma è latore di un sistema di valenze, che regolano la sua occorrenza nella struttura semantica degli enunciati in cui potrebbe essere applicato: ogni termine involupa dentro di sé istruzioni enciclopediche che lo mostrano come una sorta di sintesi di testi virtuali in cui esso potrà di fatto apparire⁷. Così |finestra| può essere accompagnato da |muro|, |parete|, |strada|, |fondo|, ecc.: in questo caso il dominio semiotico di riferimento è quello del varco che permette di illuminare gli interni delle abitazioni e di affacciarsi verso l'esterno. Si tratta di un campo semiotico differente da quello in cui il lemma |finestra| appare in un insieme di selezioni contestuali come |monitor|, |mouse|, |computer|, |utente|, ecc. – dove è la funzione informatica di interfaccia grafica ad essere presa in considerazione. L'algoritmo LSI (*Latent Semantic Indexing*) è in grado di sviluppare questo tipo di analisi linguistica, e di associare correttamente la query ai

⁷ Eco definisce ogni lemma come un *testo incoativo*: cfr. U. Eco, *Lector in fabula*, Bompiani, 1979, Milano, cap. 1.

contesti di applicazione dei termini impiegati – anche qualora essi non appaiano esplicitamente nel testo indicizzato⁸. Questo dispositivo permette di offrire agli avvocati due servizi di particolare interesse. Il primo consiste nell'introduzione di un criterio di ranking tra i principi di diritto che corrispondono alla stringa di ricerca, oltre a quello del numero di sentenze associato. Il secondo coincide con il suggerimento di varianti lessicali più efficaci della stringa di ricerca. In altre parole, autorizza l'elaborazione di un "forse cercavi" semantico e non solo ortografico: nel nostro consueto esempio architettonico, suggerisce l'applicazione di |veduta| al posto di |finestra|. Il ricorso all'algoritmo LSI introduce un meccanismo di ottimizzazione del parametro di *precision* nella compilazione della query stessa, senza rinunciare in alcuna misura all'estensione più completa del *recall*.

In terzo luogo l'architettura di *MS* interviene anche nell'organizzazione dell'elenco di sentenze che sono associate a ciascun testo normativo. Il criterio di classificazione deve collezionare le massime che vengono elencate secondo i *casi di applicazione* del principio di diritto citato, al fine di guidare la consultazione del repertorio informativo secondo una struttura logica capace di esporre le articolazioni interne alla giurisprudenza relativa al testo normativo. Il modello dell'ordinamento proviene dalle note dei *Codici Commentati* e delle *Rassegne di Giurisprudenza* che vengono pubblicate dagli editori specialistici⁹.

Il dispositivo di ricerca manifesta un comportamento intelligente, pur senza *pensare*. Tutte le azioni eseguite dal software obbediscono alle regole che governano i sistemi di intertestualità¹⁰ linguistica: la giurisprudenza è una tradizione *scritta* di interpretazione e di applicazione dei *testi* normativi, organizzata secondo una tassonomia localizzabile dai procedimenti algoritmici di *clustering*. L'iterazione ridonda le formule linguistiche dalle norme alle sentenze che li citano, alle sentenze che citano sentenze precedenti, fino alle perorazioni degli avvocati che riportano massime e sentenze. Il meccanismo di analisi e di classificazione si muove di conseguenza su una infrastruttura di *sapere scritto*, che conserva la formulazione degli enunciati e accresce la loro reperibilità statistica ad ogni nuova iterazione. L'elaborazione dell'ipotesi si sviluppa di conseguenza in una dimensione di verità pubblica, e la sua spiegazione non richiede di invocare l'attivazione di facoltà cerebrali di nessun tipo. Se questo aspetto dell'indagine può offrire un modello esplicativo anche per l'intuizione umana – ciò che di umano manca all'intelligenza del dispositivo di ricerca è il ruolo rivestito dalla responsabilità, dall'esperienza, dall'apprendimento nella costruzione stessa della verità – sono l'*ethos* del pensiero, la decisione, il progetto interpretativo della società e del diritto, che dovrebbero fondare il *valore* dell'ipotesi.

⁸ Per una comprensione dell'applicazione dell'algoritmo LSI ai motori di ricerca, cfr. C. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval*, 2006 Cambridge UP, cap. 18, pag. 277-87.

⁹ Se consideriamo per esempio il Codice Civile annotato da Cendon e Baldassarre, l'indice del commento propone la seguente articolazione per i consueti articoli concernenti le luci e le vedute: Articolo 900: 1. veduta, 2. servitù di veduta, 3. aperture lucifere, 4. giurisdizione del giudice ordinario, 5. fondo del vicino; Articolo 901:1. luci: caratteristiche; Articolo 903: 1. aperture luce irregolare. Cfr. *Codice civile annotato con la giurisprudenza*, a cura di P. Cendon e A. Baldassari, Wolters Kluwer, Milano, 2007.

¹⁰ Assumiamo la nozione di *intertestualità* nel senso originario che esso manifesta nelle opere di M. Bachtin.