

Introduzione

Linguaggio e computer

Il linguaggio umano (o *linguaggio naturale*) è lo strumento primario più immediato e duttile di cui disponiamo per creare e comunicare contenuti complessi sempre nuovi. Negli ultimi decenni, il computer si è imposto come la “macchina intelligente” per eccellenza, capace di memorizzare ed elaborare quantità enormi di dati strutturati e di renderli disponibili come informazioni. Ciononostante, le sue capacità “comunicative” sono rimaste sorprendentemente rudimentali, sostanzialmente affidate alla mediazione di strutture logiche predefinite, menu preconfezionati e linguaggi artificiali più o meno ad hoc. L’obiettivo centrale della linguistica computazionale (LC) è quello di sviluppare *modelli computazionali* della lingua, cioè modelli del funzionamento del linguaggio naturale che possano essere tradotti in *programmi* eseguibili dal calcolatore e che consentano a quest’ultimo di acquisire le competenze necessarie per comunicare *direttamente* nella nostra lingua.

Allo stato attuale delle nostre conoscenze, questo obiettivo può essere perseguito a più livelli di astrazione. Immaginiamo di pronunciare una richiesta di informazioni davanti a un microfono collegato al *riconoscitore vocale* di un computer. In prima battuta, possiamo ottenere in risposta la trascrizione ortografica automatica delle parole da noi pronunciate. A un secondo livello, il computer potrebbe analizzare il contenuto della richiesta trascritta, associandolo a una rappresentazione semantica appropriata. Un calcolatore ancora più “evoluta” potrebbe rendere disponibili le informazioni oggetto della richiesta, dopo averle reperite da una fonte di dati accessibile, e così via. Ciascuno di questi compiti richiede l’acquisizione di competenze molto specifiche, da tradurre in *tecnologie linguistiche* accessibili al calcolatore. È chiaro, tuttavia, che la capacità di comunicare con l’ambiente esterno in modo intelligente passa attraverso l’*integrazione*

di queste tecnologie in *sistemi* funzionalmente complessi, in grado di interpretare l'informazione in ingresso, elaborarla e rispondere con altra informazione pertinente. Sono sistemi con queste caratteristiche che si nascondono dietro un *risponditore automatico*, un *traduttore automatico* o un *classificatore di pagine web*. In un futuro non troppo lontano, potremmo sorprenderci a interagire con una macchina intelligente che percepisce e ragiona sulla realtà come un essere umano!

Il livello di astrazione appropriato, gli obiettivi intermedi di ricerca, i metodi e i modelli di spiegazione della LC possono variare a seconda del punto di vista particolare da cui guardiamo al rapporto tra linguaggio e calcolatore. Ad esempio, possiamo chiederci *cosa il linguaggio naturale può fare per il computer*. In che misura le attuali potenzialità dei computer possono essere migliorate ed estese dotando queste macchine di capacità linguistiche evolute? Viviamo, infatti, in quella che è stata spesso definita come la *società della conoscenza*, dove studenti, professionisti e comuni cittadini si trovano nella necessità quotidiana di recuperare, selezionare e gestire grandi quantità di informazioni contenute nelle miriadi di testi disponibili sul *web*, nelle biblioteche digitali o sui computer di casa o di lavoro. La facilità di accesso a queste informazioni e la capacità di distinguere le parti rilevanti da quelle inutili sono condizioni essenziali per garantire che si realizzino in pieno le potenzialità della rivoluzione informatica in corso. In questa prospettiva, la LC promette di espandere enormemente le capacità del computer, rendendolo capace di comprendere la struttura e il contenuto dei nostri testi e di interagire col mondo esterno con la naturalezza e l'immediatezza del linguaggio naturale.

È possibile però rovesciare il rapporto tra computer e linguaggio umano e chiederci *cosa il primo può fare per il secondo*: ovvero, in che modo il computer possa essere messo al servizio dell'indagine linguistica. A differenza della precedente, questa è una prospettiva squisitamente legata alla *linguistica* e alle *scienze umane*, il cui obiettivo finale è una migliore comprensione del modo in cui il linguaggio funziona e si rapporta alle altre facoltà cognitive. Da questo punto di vista, il ruolo dei metodi matematici e informatici è di contribuire a definire modelli che spieghino proprietà e problemi linguistici interessanti. Le domande che si pone il linguista computazionale sono dunque le stesse del linguista tradizionale: come funziona il linguaggio, come viene appreso, come cambia ecc., e la dimensione innovativa è data dal *modo* in cui il computer aiuta lo studioso a impostare questi problemi e a verificarne le soluzioni.

Distinguere queste due prospettive non significa però vederle in maniera indipendente o antagonista. La ricerca linguistica degli ulti-

mi decenni ha consentito lo sviluppo di programmi applicativi che permettono una migliore gestione dell'informazione, proprio grazie a una più profonda comprensione del linguaggio. Allo stesso modo, metodi matematici e informatici nati e sviluppati per scopi autonomi si sono spesso rivelati assai utili per innovare le metodologie di rappresentazione e le analisi linguistiche. Le due direzioni nel rapporto tra linguaggio e computer sono dunque profondamente compenstrate e in osmosi continua. La LC è, infatti, un ambito di ricerca e di lavoro tipicamente *interdisciplinare*. Da un lato, essa si intreccia con l'*informatica*, sia teorica che applicata, e con l'*ingegneria*, come testimonia l'uso del termine *ingegneria delle lingue* (o *del linguaggio*, inglese *natural language engineering*) per qualificare le attività rivolte allo sviluppo di applicazioni informatiche basate sulle tecnologie del linguaggio. Dall'altro lato, invece, la LC dialoga con la *linguistica*, le *scienze cognitive*, la *psicologia*, la *filosofia* e le *scienze umane* in generale, con le quali condivide l'obiettivo di indagare la struttura, il funzionamento e l'uso del linguaggio e il suo rapporto con le altre facoltà cognitive dell'uomo. In conseguenza di questo carattere interdisciplinare, linguisti, informatici, ingegneri, logici e umanisti, ma anche neuro-fisiologi, bio-ingegneri e filosofi della mente si trovano sempre più spesso a lavorare insieme su temi comuni di ricerca. Il linguista computazionale "ideale" dovrebbe essere in grado di riassumere in sé in maniera equilibrata le caratteristiche più salienti di tutte queste anime. Naturalmente, a seconda dell'obiettivo – più o meno applicativo – della ricerca, il peso relativo di ciascun tipo di competenza può cambiare. Il successo finale, tuttavia, dipende dalla profondità della loro interazione.

Alla luce di queste considerazioni, è nostro convincimento che non solo una sintesi tra le varie anime della LC è possibile, ma anche *strettamente necessaria*. Troppo spesso, invece, assistiamo a scollamenti in entrambe le direzioni. In alcuni casi, l'ingegneria delle lingue tende a procedere in maniera indipendente dalla ricerca linguistica, sulla spinta di esigenze di competizione tecnologica e di mercato. Viceversa, non di rado il computer è visto come un semplice strumento per potenziare il modo tradizionale di fare ricerca sul linguaggio. In realtà, la LC è anche un modo *nuovo* di guardare al linguaggio, di studiarne la struttura e l'uso, di ripensarne l'organizzazione e le dinamiche. Come in tutte le innovazioni tecnologiche, l'applicazione del computer al linguaggio sta progressivamente mutando il modo in cui immaginiamo, conosciamo e usiamo quest'ultimo. Al tempo stesso, la complessità del fenomeno è tale che lo sviluppo di applicazioni informatiche dotate di intelligenza linguistica non può prescindere dai

problemi che l'organizzazione e il funzionamento del linguaggio pongono.

La linguistica computazionale tra ieri e oggi

È agli inizi della seconda metà del Novecento che può essere ricondotta la nascita della LC come disciplina di frontiera sia rispetto al mondo umanistico, che rispetto alle applicazioni più centrali delle scienze dell'informazione che stavano sviluppandosi in quegli stessi anni. Le sue radici affondano in due distinti paradigmi di ricerca. Da un lato, gli anni cinquanta e sessanta vedono le prime applicazioni del calcolatore allo studio dei testi filosofici e letterari, con il lavoro pionieristico di padre Roberto Busa S. J. che, presso il Centro per l'automazione dell'analisi linguistica di Gallarate, realizza il primo corpus elettronico delle opere di Tommaso D'Aquino (circa 10 milioni di parole, una cifra enorme per le capacità dei computer del tempo) e un programma per la loro esplorazione attraverso *concordanze* (cfr. PAR. 7.2). Lo sviluppo di programmi per lo spoglio elettronico dei testi, il calcolo della frequenza delle parole, la compilazione di indici e concordanze, la creazione di repertori lessicali elettronici (i cosiddetti *dizionari macchina*) sono tra i primi risultati della disciplina.

Il secondo paradigma di riferimento è rappresentato invece dall'applicazione di *metodi formali* all'analisi del linguaggio. Il 1957, data di nascita della grammatica generativa e della "rivoluzione linguistica" di Noam Chomsky con la pubblicazione di *Syntactic Structures*, segna anche l'inizio di un'intensa attività di indagine, con l'ausilio di grammatiche formali, sulle proprietà del linguaggio naturale. A sua volta, questo lavoro è profondamente intrecciato con le ricerche coeve nel settore dell'Intelligenza Artificiale (IA). È all'interno di questa tradizione che si sviluppa il settore del *Natural Language Processing* (NLP) o Trattamento Automatico del Linguaggio Naturale (TAL), con i primi programmi per l'analisi sintattica e l'interpretazione semantica automatica, e applicazioni ormai classiche come la *traduzione automatica* e le cosiddette *interfacce uomo-macchina* in linguaggio naturale.

Nei suoi primi decenni (anni sessanta e settanta), il TAL è stato fortemente influenzato dalla metodologia *razionalista e simbolica* allora dominante in IA e condivisa dalla linguistica generativa. Ciò ha significato essenzialmente il predominio di metodi logico-deduttivi e algebrici, a discapito di quelli induttivi e statistico-matematici. Il modello più influente della conoscenza umana, in questa fase, è quello di un sistema di *regole* per la manipolazione di strutture di simboli.

Nel caso del linguaggio, queste regole riguardano la *conoscenza* che i parlanti hanno di una lingua, e che determina la loro possibilità di giudicare se una certa frase è grammaticalmente corretta. Nella prospettiva chomskiana, infatti, fondamentale è la distinzione tra *competenza* ed *esecuzione*, ovvero tra la *conoscenza* del linguaggio che hanno i parlanti e l'*uso* che di essa fanno nelle situazioni comunicative concrete, per comprendere una frase o descrivere un evento. Le informazioni quantitative, secondo Chomsky, non hanno nessun ruolo nello studio della competenza: questa consiste infatti in un sistema di regole strutturate e formulate in modo del tutto indipendente dall'uso. La distribuzione statistica delle costruzioni linguistiche è al più rilevante nella sfera dell'esecuzione, e resta quindi esclusa dall'ambito di indagine proprio della grammatica generativa. Come conseguenza diretta di questo approccio epistemologico, sviluppare un modello computazionale del linguaggio ha voluto dire per lungo tempo scrivere *sistemi di regole linguistiche* interpretabili dal calcolatore. La mancanza di attenzione verso i problemi dell'uso linguistico ha portato, in molti casi, alla creazione di *modelli giocattolo* (*toy models*), ovvero applicazioni che, seppure in grado di analizzare alcune costruzioni linguistiche non banali, non si sono dimostrate capaci, anche a causa dei limiti tecnologici dell'epoca, di uscire dai laboratori e operare in contesti reali, con vocabolari di alcune migliaia di parole e su frasi di lunghezza media.

Parallelamente al diffondersi della grammatica generativa, ha continuato a svilupparsi, soprattutto in area anglosassone, una tradizione di ricerca linguistica ancorata a una metodologia di derivazione *empirista*, che fonda l'indagine linguistica sulla raccolta e l'analisi di *corpora* (cfr. CAP. 1), ovvero raccolte di grandi quantità di testi appartenenti a una certa varietà del linguaggio. La *linguistica dei corpora* (*corpus linguistics*) fa largo uso di strumenti di analisi quantitativa e statistica per esplorare le regolarità linguistiche che emergono dai testi e che vengono a costituire la base per la descrizione della struttura del linguaggio. A differenza della grammatica generativa, la descrizione del linguaggio è giudicata inseparabile dall'analisi del suo uso. I metodi pionieristici di indagine quantitativa, originariamente applicati all'analisi dei testi letterari, sono stati dunque rapidamente estesi a testi di linguaggio comune, come corpora di giornali, trascrizioni di parlato ecc. Questo tipo di indagine ha fortemente beneficiato della capacità dei computer di gestire ed elaborare quantità di dati sempre più estese. Nel 1964 compare il *Brown Corpus* (lungo circa 1 milione di parole), realizzato da W. Nelson Francis e Henry Kucera alla Brown University negli Stati Uniti. Questo è il primo corpus elettronico pro-

gettato e costruito per lo studio di una particolare varietà linguistica contemporanea, nel caso specifico l'inglese americano degli anni sessanta. Uno dei risultati di questa linea di ricerca è stato quello di favorire l'applicazione al linguaggio di una serie di metodi statistici tuttora fondamentali in LC (cfr. CAPP. 5 e 6).

Per molti anni, la LC di ispirazione razionalista e quella di tradizione empirista hanno proceduto in maniera parallela, con pochi e sporadici contatti. Il superamento di tale dicotomia è stato reso possibile grazie al radicale mutamento di paradigma avvenuto nella LC a partire dalla seconda metà degli anni ottanta, caratterizzato dal diffondersi e poi dal netto prevalere di un approccio neoempirista al TAL. Questo cambiamento si è concretizzato nella diffusione dei metodi statistico-quantitativi e nella rinnovata centralità dei dati testuali come base essenziale per l'esplorazione del linguaggio. La dimensione dei corpora è andata sempre più crescendo, e la loro disponibilità è diventata una variabile fondamentale in ogni fase di sviluppo e valutazione degli strumenti per il TAL. Al tempo stesso, la crescita del web, con la sua grande quantità di materiale testuale multiforme e in continua evoluzione, ha reso necessario lo sviluppo di tecnologie linguistiche in grado di confrontarsi non più con esempi da laboratorio, ma con testi *reali*. Questo ha a sua volta contribuito al diffondersi di tecniche di elaborazione linguistica più "robuste" di quelle simboliche tradizionali, tecniche in grado di affrontare la variabilità lessicale e strutturale dei testi, e la continua violazione delle regole che ha luogo in molte manifestazioni dell'uso linguistico: prima fra tutte, la lingua parlata.

La crescente disponibilità di corpora testuali ha indubbiamente facilitato l'innovazione metodologica in LC, fornendo i dati linguistici necessari per un uso *intensivo* dei metodi statistici. L'emergere di linguaggi standard di *marcatatura* del testo come XML (cfr. CAPP. 2 e 3), inoltre, ha migliorato il processo di creazione, uso e interscambio dei corpora. Al tempo stesso, ha reso possibile l'annotazione linguistica del testo su ampia scala, che consente di rappresentare in maniera esplicita e dettagliata la struttura e l'organizzazione linguistica del testo stesso (cfr. CAPP. 3 e 8), offrendo l'occasione per una riflessione rigorosa sulla natura dei dati annotati e sui modelli interpretativi più adatti per darne conto.

Parallelamente, anche in linguistica teorica si è incominciato a prestare nuova attenzione al dato statistico, e al ruolo dei modelli probabilistici nella descrizione linguistica e cognitiva in generale. Ad esempio, nelle scienze cognitive si è fatta sempre più strada la con-

vinzione che la mente umana non sia soltanto capace di elaborare regole, ma anche di tener traccia di regolarità statistiche nell'ambiente che la circonda.

Cos'è dunque la LC oggi? Sicuramente è una disciplina scientifica matura, che in 50 anni è riuscita a conquistare una posizione di indiscussa centralità nel panorama scientifico nazionale e internazionale, e in cui sono attivi ormai innumerevoli gruppi di ricerca. La Association for Computational Linguistics (ACL) organizza una conferenza annuale a cui partecipano ricercatori di tutto il mondo. A questa si affianca come importanza l'International Conference on Computational Linguistics (COLING) che si tiene con cadenza biennale. In Italia, Antonio Zampolli (1937-2003), allievo di padre Busa, ha fondato e diretto per lunghi anni l'Istituto di Linguistica Computazionale del Consiglio Nazionale delle Ricerche (CNR) e ha occupato la prima cattedra di Linguistica matematica creata presso l'Università di Pisa nel 1977 (Zampolli *et al.*, 2003). Attualmente, centri di ricerca pubblici e privati sono attivi su tutto il territorio nazionale, e la LC e l'informatica applicata alle scienze umane sono diventate insegnamenti fondamentali in molte università, che offrono spesso percorsi di studio interdisciplinari umanistici e informatici.

La storia della LC ne ha anche determinato il carattere complesso sia nell'articolazione interna sia nei rapporti con altre discipline, che comprendono l'analisi informatica dei testi letterari (il cosiddetto *literary computing*), le scienze cognitive, l'IA e una miriade di settori applicativi legati all'ingegneria delle lingue. Sul versante della ricerca, i modelli statistici hanno ancora un ruolo dominante. Dopo averne esplorato non solo le grandi potenzialità, ma anche i limiti intrinseci, il settore va tuttavia sempre più orientandosi verso un loro uso più consapevole, attraverso lo sviluppo di modelli e strumenti *ibridi*, che incorporano sia conoscenza rappresentata come regole, sia moduli statistici. Sul piano tecnologico, le applicazioni basate sul TAL incominciano a diffondersi nella vita di tutti i giorni. È sempre più comune imbattersi in siti web e call center che usano tecnologie e strumenti linguistici, spesso invisibili all'utente comune, ma tanto più efficaci quanto meno ne è percepita la presenza. Ciononostante, queste tecnologie sono ancora limitate se confrontate alle competenze linguistiche di un parlante. I limiti non sono solo tecnologici, ma riguardano anche la nostra comprensione, ancora parziale, di come il linguaggio funziona. Anche in questo sta il fascino della LC: una disciplina in cui il progresso tecnologico si deve sposare con la nostra capacità di penetrare i misteri del linguaggio e della mente umana.

Di cosa parla questo libro

Questo libro è il primo di due volumi dedicati al rapporto tra linguaggio naturale e calcolatore. I temi che abbiamo selezionato per questo volume definiscono lo *strumentario* di base per il linguista computazionale e in genere per chiunque voglia affrontare in maniera proficua l'argomento partendo da un interesse primario per il *testo*, la sua struttura e il suo contenuto. Nella prima parte, *Il dato testuale e la sua rappresentazione*, il lettore troverà indicazioni utili riguardo ai criteri per *la selezione* e *l'uso* del materiale testuale primario (CAP. 1) e alle *metodologie* di rappresentazione e codifica dell'informazione linguistico-testuale (CAPP. 2 e 3). L'accesso del computer al testo non è affatto immediato: anche unità banali come il titolo, il paragrafo o la singola parola sono in gran parte implicite e nascoste. Codificare un testo digitale significa, in sostanza, tracciare una prima, grossolana mappa virtuale che consenta al calcolatore di muoversi nel labirinto del testo con un minimo di consapevolezza di dove si trovi. Nella seconda parte, *Metodi e tecniche di analisi testuale*, la complessità linguistica del testo comincia a emergere agli occhi del computer, attraverso un paziente lavoro di ricostruzione a partire dal flusso ininterrotto di *bit*. Si parte dall'identificazione di unità lessicali semplici e complesse attraverso l'uso di espressioni regolari (CAP. 4), per proseguire con la descrizione della loro distribuzione nel testo (CAP. 5) e con l'individuazione di relazioni strutturali tra parole mediante modelli probabilistici (CAP. 6). La terza e ultima parte del libro, *Esplorazione e annotazione del testo*, infine, riporta l'enfasi sul lavoro del linguista e sulla sua capacità di estrarre ulteriori informazioni dal testo attraverso percorsi mirati di ricerca, condotti su materiale sia grezzo che *annotato* (CAP. 7). È spesso assai utile, infatti, effettuare queste ricerche su testi annotati che diventano così uno scrigno di informazioni linguistiche preziose, automaticamente accessibili a chiunque voglia farne uso per ulteriori ricerche. A questa delicata fase di annotazione linguistica è dedicato il CAP. 8. Infine, l'ultimo capitolo getta un ponte verso il secondo volume accennando al problema dello sviluppo di *modelli espliciti* della conoscenza linguistica, come ad esempio gli analizzatori morfologici.

Alcune avvertenze preliminari si rendono necessarie. Questo libro non è né un'introduzione alla linguistica, né una rassegna delle applicazioni tecnologiche della LC. Per rendere tuttavia il volume ugualmente leggibile sia per linguisti e umanisti che per informatici e ingegneri, abbiamo cercato di limitare per quanto possibile le conoscenze presupposte, sul versante linguistico come su quello matematico e in-

formatico. A tale scopo, schede tecniche di approfondimento sono disponibili sul sito <http://www.carocci.it>, dove il lettore potrà anche trovare utili esercizi raggruppati per capitoli. Per le stesse ragioni, abbiamo evitato di trasformare il volume in una *summa teorica* del dibattito scientifico attuale. Al contrario, ci siamo sforzati di raggiungere un equilibrio tra il *sapere* e il *fare*: nozioni di carattere teorico sono sempre accompagnate dalla presentazione di tecniche e indicazioni metodologiche utili per poter analizzare un testo al calcolatore, per scopi di ricerca o applicativi.

Tra le innumerevoli sfaccettature del linguaggio, particolare importanza assume la distinzione tra lingua scritta e parlata, le due principali *modalità* attraverso cui passa la comunicazione linguistica. Negli ultimi vent'anni, il settore delle *tecnologie del parlato* ha affrontato con grande successo i problemi relativi alla *sintesi* e al *riconoscimento* vocali da parte del calcolatore. Assai più recente, in ambito computazionale, è invece l'interesse per il rapporto tra *comunicazione linguistica*, *gesti* ed *espressioni facciali* e, più in generale, per l'integrazione tra linguaggio e percezione sensoriale, nella direzione di un approccio *multimodale* al TAL.

Da parte nostra, abbiamo deciso di focalizzare l'attenzione su quegli aspetti *centrali* del linguaggio umano che prescindono dalle particolari modalità attraverso cui passa la comunicazione linguistica. A questa ragione va ricondotta la decisione di scegliere, come elemento cardine del nostro percorso espositivo, il *testo*, inteso come qualsiasi prodotto dell'attività linguistica che sia *elaborato o trascritto in forma di caratteri*. Questa definizione comprende libri, manoscritti, lettere, articoli ma anche dialoghi, trasmissioni radiofoniche o conversazioni telefoniche nella misura in cui questi ultimi sono trascritti in un qualche sistema grafico di caratteri. Il testo oggetto delle nostre analisi non è dunque limitato alle forme di produzione linguistica progettate per la modalità scritta, come libri o giornali. Pertanto, i metodi e gli strumenti che verranno illustrati su testi scritti si applicano altrettanto bene al linguaggio parlato trascritto. Ad esempio, alcuni dei metodi quantitativi illustrati nei CAPP. 5 e 6 di questo volume sono stati sviluppati al servizio delle tecnologie per il riconoscimento del parlato, ma la loro rilevanza dal punto di vista dello studio della lingua è indipendente dalle specificità del segnale acustico.

Il web e l'evoluzione telematica hanno inoltre reso molto più fluido il confine tra lingua scritta e lingua parlata. La posta elettronica, le chat, i blog e gli SMS rappresentano alcuni dei nuovi stili di comunicazione in cui lo strumento elettronico e la forma dell'interazione fanno sì che il testo scritto contenga in realtà tratti lessicali, sintattici ed

espressivi tipici della lingua parlata. Adottando una nozione *aperta* di testo, gli stessi metodi e strumenti computazionali possono dunque essere applicati a tutta una serie di prodotti del linguaggio rappresentativi dell'intera gamma delle modalità comunicative, delle varietà linguistiche e dei registri dell'italiano contemporaneo. La centralità che il testo assume nella nostra trattazione ci ha portato inoltre a concentrarci prevalentemente su aspetti d'*uso* del linguaggio. È nostro convincimento, tuttavia, che lo studio del linguaggio mediante il calcolatore possa contribuire a ricomporre e superare in una nuova sintesi la dicotomia tra conoscenza e uso, competenza ed esecuzione linguistica, rappresentazione e computazione, gettando nuova luce sul rapporto tra facoltà del linguaggio e cognizione.

Concludiamo con una nota terminologica. La difficoltà di reperire traduzioni italiane dei tecnicismi del settore ha in genere stimolato la nostra creatività lessicale. Tutte le volte che ci è stato possibile, abbiamo scelto di tradurre il termine tecnico in un modo che suonasse naturale e ragionevole per un parlante italiano e che richiamasse alla mente, al tempo stesso, la sua fonte originale (menzionata alla prima apparizione della traduzione nel testo) per non disorientare troppo il lettore che abbia già una qualche dimestichezza con la letteratura. È possibile che, nonostante i nostri sforzi, le scelte compiute non siano sempre state le più felici e che qualche lettore purista possa storcere il naso di fronte ad alcuni inevitabili prestiti e calchi. Abbiamo preferito questo rischio, tuttavia, a quello di rendere il testo meno comprensibile per il neofita, nascondendo concetti di base dietro etichette generiche e poco comprensibili.