



DALL'INDICIZZAZIONE ALL'ERMENEUTICA TESTUALE

FILOSOFIA DEL LINGUAGGIO E LINGUISTICA COMPUTAZIONALE

di Savina Raynaud

Una vasta gamma di discorsi compiuti, di diversa estensione, cade sotto il concetto di testo, o dei suoi quasi-sinonimi messaggio e documento. La linguistica testuale ha esteso l'unità d'analisi dagli enunciati ai testi e ha enunciato le relative condizioni di buona formazione. Entro l'orizzonte testuale è possibile sceverare, con il contributo della filosofia del linguaggio, i costituenti della semiosi verbale, le regole delle relative codifiche e il divenire nel tempo dei supporti fisici dell'informazione e della comunicazione. Sulla base di queste distinzioni viene spiegato il significato dell'opposizione analogico-digitale. In merito alla linguistica computazionale si tratteggia infine da un lato la storia dell'impresa pionieristica di P. Roberto Busa e della realizzazione dell'*Index Thomisticus*, dall'altro l'attuale continuazione di quel progetto e il costituirsi di un' articolata comunità scientifica. Il trattamento automatico del linguaggio naturale affina e irrobustisce il compito sempre attuale e delicato dell'ermeneutica testuale.

A wide range of complete discourses, of different extensions, fall either into the category of the text or one of its near synonyms, such as messages or documents. Text linguistics has gone beyond the unit of sentences in analyzing texts; and has stated the conditions of well-formedness of texts. Within the textual framework, thanks to the philosophy of language, it is possible to distinguish the constituents of verbal semiosis, the rules of encoding and the periodic transformations of physical media for encoding and transporting information and communication. On the basis of these premises it is possible to explain the meaning of the analogue-digital opposition. As far as computational linguistics is concerned, we provide a short treatment of the pioneering work of Fr. Roberto Busa and of the *Index Thomisticus* on the one side, and on the other, we sketch the present development of this project and the constitution of a lively scientific community of computational linguists. Automatic natural language treatment refines and strengthens the delicate task of textual hermeneutics.

1. DALLA FENOMENOLOGIA ALL'IDENTIFICAZIONE:

ZU DEN SACHEN SELBST

A più titoli ci si può rivolgere a un testo e alla relativa nozione: in sordine ad altro o in prospettiva *autotelica*, cioè con una finalità che al testo (e al concetto di testo) punta e li vuole sostenere.

Se è tipico della filosofia interrogarsi a volta a volta su questioni specifiche mantenendo sullo sfondo l'intero, è tipico delle scienze concentrarsi su domini determinati. Proverò a rispondere al quesito assegnato – quale *fenomenologia* del testo proporre all'intersezione in cui si collocherebbe *l'informatica umanistica*? – coniugando prospettive tipiche della filosofia del linguaggio con alcuni risultati conseguiti in linguistica computazionale, alla luce di progetti in corso.

Comincerò col rilevare che interrogarsi su *che cos'è un testo* vuol dire disporsi ad affrontare un compito per molti versi ben più impegnativo di quanto non sia fermarsi a fornire una *fenomenologia* del testo. Socrate, insomma, sospinge avanti quesiti che possono, sì, trarre beneficio dall'invito a non trascurare ciò che appare – i fenomeni, appunto –, a soffermarsi anzi su ciò che è manifesto invece di precipitarsi a inferire trascurando o travisando i dati. Chi indaga in sintonia con Socrate può apprezzare da un lato, con il primo Husserl, il motto *zu den Sachen selbst*¹ e di conseguenza l'indicazione di metodo che ne scaturisce – *wir sind echte Positivisten*² –. Ma mal sopporterebbe la consegna dell'*epoché*, una sospensione di giudizio non indefinitamente sostenibile. Alla domanda radicale – *che cos'è?* – va data la migliore tra le risposte possibili, la meno evasiva, la più approssimata anche se non ultimamente risolutiva, comunque.

Che cos'è dunque un testo? È possibile fare una fenomenologia del concetto senza osservarne molti? O il concetto si astrae e lo si illustra, a partire da una molteplice esperienza di fruizioni testuali, oltre

¹ *Ricerche logiche, Introduzione § 2: alle cose stesse.*

² *Idee per una fenomenologia pura e per una filosofia fenomenologica*, libro I, sezione I, cap. II, § 20: *Se positivismo è la fondazione assolutamente libera da pregiudizi di tutte le scienze sul positivo, cioè su quello che si afferra originalmente, allora siamo noi i veri positivisti.*

che di produzioni testuali, in quanto autori? Come inserire il concetto di *testo* nella rete concettuale?

2. L'OGGETTO TESTO E LA SUA IDENTITÀ

Da una duplice angolatura si può dire che il testo in quanto tale non è immediatamente un osservabile, ma è un costruito, non è attinto da un'astrazione elementare, ma è il derivato di una sequenza di procedure astrattive e inferenziali: sia dal punto di vista della formazione del termine, sia da quello dell'epistemologia e della storia delle scienze del linguaggio.

Se infatti andiamo all'etimologia del latino *textus* a cui tutte le rese traduttive nelle principali lingue occidentali riconducono (it. *testo*, sp. *texto*, fr. *texte*, ingl. *text*, ted. *Text*, ma anche ceco *text* ecc.), dobbiamo prendere atto che si tratta di un participio sostantivato da *texere*, con valore traslato, dall'attività propriamente tessile a quella del comporre parole secondo un intreccio appropriato e in base a una struttura che prevede un inizio e una fine, nonché uno sviluppo intermedio.

È d'altra parte solo una tappa del secondo Novecento in linguistica generale l'elaborazione di una linguistica testuale³ che, quale unità base della propria analisi e generazione modellistica, poneva il testo, entità irriducibile ai suoi componenti frastici e limite superiore alla sequenza ben formata di questi⁴.

Appare, direi, la retorica⁵, il sapere antico più prossimo a dar conto della gradualità di costituzione del *dictum – inventio, dispositio, elocutio, memoria, actio* –, con una priorità marcata riservata al parlato e al

³ Cfr. M.E. Conte (a cura), *La linguistica testuale*, Feltrinelli, Milano 1977; C. Andorno, *Linguistica testuale. Un'introduzione*, Carocci, Roma 2003.

⁴ Cfr. L. Rocci, *La testualità*, in G. Bettetini, S. Cigada, S. Raynaud, E. Rigotti (a cura), *Semiotica II. Configurazione disciplinare e questioni contemporanee*, La Scuola, Brescia 2003, pp. 257-319, con ampia bibliografia.

⁵ Cfr. C. Marmo, S. Bonfiglioli (a cura), *Retorica e scienze del linguaggio. Teorie e pratiche dell'argomentazione e della persuasione*, Atti del X congresso nazionale Rimini 2003, Pubblicazioni della Società di Filosofia del Linguaggio 01, Aracne, Roma 2005.

perlocutorio ⁶ di questo. Ma già la *Poetica* aristotelica (20, 28-30) coglie con formidabile precocità l'indipendenza dell'unità testuale da criteri meramente quantitativi, raccogliendo sotto il medesimo titolo di *logos* l'Iliade e la definizione di uomo, unitarie entrambe, l'una per congiunzione, l'altra perché dotata di un unico significato.

Troppo lungamente, invece, la prospettiva metalinguistica è stata esercitata sulle unità minime ⁷ e al più su unità superiori del rango sintattico di frasi e periodi, ma non su unità testuali.

Troppo spesso si è poi ristretto arbitrariamente il concetto di testo all'ambito dello scritto, mentre nel Novecento c'è chi ha esplicitato l'ammissibilità di una sinonimia fra *testo* e *messaggio* (che può essere orale o scritto) ⁸ o più recentemente ⁹ ha tematizzato la rilevanza di una nozione centrale per le tecnologie dell'informazione e della comunicazione (d'ora in poi ICT), quella di *documento*.

Le forme interne ¹⁰ di queste tre nozioni, del resto, dicono di una complementarità di diritto dei rispettivi tratti salienti nei relativi referenti (nei testi, cioè, o messaggi, o documenti) che, se di fatto non è esibita da ciascun termine, non è certo esclusa, nemmeno di fatto, negli oggetti denominati: il *messaggio* (di nuovo dal latino *missum* di *mittere*) rimanda all'invio a destinazione del tessuto verbale, mentre la funzione indicativa, istruttiva del messaggio medesimo è esaltata nel *documentum* (da *docere*).

Se tuttavia questi e molti altri ruoli possono essere ascritti ai testi, molto spesso sono le prospettive d'uso o gli ambiti di fruizione a determinarne lo statuto: ora di opere (d'ingegno, di lettere), ora di atti (amministrativi, giuridici), ora di riferimenti normativi, di testimonianze (lettere, comunicazioni personali), di fonti (imprescindibili per la ricostruzione storica) e così via.

⁶ Nell'accezione di Austin.

⁷ Sul lessico e sulle parti del discorso, sui morfemi variabili, addirittura su foni e fonemi, su loro tratti distintivi.

⁸ Così nel modello di Jakobson delle funzioni comunicative.

⁹ Cfr. M. Ferraris, *Documentalità. Perché è necessario lasciar tracce*, Laterza, Roma-Bari 2009.

¹⁰ Cioè le immagini, le rappresentazioni collaterali al significato o a esso ausiliarie, suscitate dai costituenti delle espressioni linguistiche.

Il prevalere dell'attenzione sui fini intesi rispetto ai mezzi testuali a volta a volta considerati ha spesso determinato una spiccata tendenza a subalternare ad altro l'attenzione riservata al *testo in quanto testo*.

Tuttavia alcune conclusioni teoriche sono state tratte, in particolare l'identificazione delle condizioni di buona formazione dei testi; sistema aperto, tuttavia articolato in alcuni criteri di base, quali continuità referenziale, coesione, coerenza, informatività¹¹.

Già, un testo, per esser tale, non deve solo avere una sua continuità di riferimenti, non deve solo essere interconnesso e privo di lacune, non contraddittorio, ma deve anche *portare informazione*, deve mettere-in-forma, in certa misura formalizzare i propri contenuti.

3. NON È SOLO QUESTIONE DI SUPPORTI

Dunque, come per il linguaggio nella sua totalità, così per un testo vale la pena domandarsi: che cos'è complessivamente? Qual è il *segreto* della sua unità, della sua interna correlazione? È un'entità fisica? O socio-culturale? O entrambe? E come dar conto di questo Giano bifronte? Troppo spesso si parla di digitalizzazione di testi, di biblioteche digitali dando per certo che il livello risolutivo sia quello dei supporti, da cartaceo a elettronico, da CD alla rete ecc. Senza nulla togliere alla rilevanza di questi passaggi – e al loro variegato *fare la differenza* – non possiamo non domandarci *che cosa sia* ciò che ne è supportato, che cosa resti di invariante e irriducibile ai diversi supporti fisici su cui anche un medesimo testo può essere *caricato*. Più ancora, come sia possibile dar conto della compresenza e del legame strutturato, dell'*interfaccia* tra supporto e supportato.

Interrogativi simili, e al contempo sollecitazioni a vedere il lato materiale dell'evoluzione culturale, sono stati alla base dell'iniziativa di allestire una mostra in coincidenza con l'XI Congresso Nazionale della Società di Filosofia del Linguaggio, ospitati entrambi dal-

¹¹ Cfr. R. De Beaugrande, W. Dressler, *Text Linguistics*, Longman, Harlow 2004, i quali aggiungono intenzionalità, accettabilità, situazionalità, intertestualità.

l'Università Cattolica nel settembre 2004, a Milano ¹². L'intento era duplice: sensibilizzare gli studiosi di semantica al coinvolgimento del mondo fisico nell'attività di significazione e sensibilizzare gli studiosi di altre aree alla necessità di includere esplicitamente il mondo dei significati – così invisibile, impalpabile, silente – nell'osservazione della scrittura e della sua storia, pena l'incomprensione della stessa ragion d'essere dello scrivere e delle sue evoluzioni, e prima ancora del parlare e del suo divenire.

Quale che sia l'avvicinarsi tra parola parlata e parola scritta, *l'alleanza indissolubile che unisce il pensiero, gli organi vocali e l'udito al linguaggio risiede in modo irrevocabile nella costituzione originaria, non ulteriormente esplicabile, della natura umana*, scriveva infatti Humboldt nel 1835, ne *La diversità delle lingue*. Eppure la ricerca linguistica, in quel secolo d'oro che fu l'Ottocento, si sarebbe orientata in misura più che preponderante ai documenti scritti, e sarebbe così venuta privilegiando quelle prospettive diacroniche d'indagine, che solo la scrittura consentiva di assumere. La lingua parlata, e dunque il suo fluire sonoro, restavano lontani dalla dignità documentale che ogni ricerca, tanto più se di impostazione storica, esigea. L'interesse per il parlato, pur presente nell'Ottocento (lingue esotiche, parlate locali, dialetti; da parte di viaggiatori, missionari o altri osservatori occasionali), non raggiunse la visibilità della corrente dominante della linguistica ufficiale ¹³.

Noi stessi tuttora tendiamo spesso – come nel caso delle considerazioni su ciò che è un testo – a neutralizzare l'opposizione, per tanti versi marcata, tra scritto e parlato.

Le nuove tecnologie favoriscono d'altra parte, insieme con una moltiplicazione del ricorso alla scrittura e alla sua trasmissione telematica, formidabili occasioni di riconversione dei due *formati*, scritto e parlato: dalla sintesi vocale dello scritto alla trascrizione automatica del parlato.

¹² Titolo della mostra: *La voce e la scrittura*. Per gli Atti del Congresso cfr. A. Frigerio, S. Raynaud (a cura), *Significare e comprendere. La semantica del linguaggio verbale*. Atti dell'XI congresso nazionale. Pubblicazioni della Società di Filosofia del Linguaggio 02 Milano 2004, Aracne, Roma 2005.

¹³ Cfr. S. Aurox, *La catégorie du parler et la linguistique*, in «Romantisme» 25-26 (1979), pp. 157-178.

Ma cerchiamo di procedere con ordine.

4. ALFABETI, IDEOGRAFIE, CODIFICHE

Se dunque puntiamo al nesso tra supporti e supportati, tra mezzi e fini, tra forme e contenuti, un livello sensibile della loro correlazione è appunto quello dell'in-formazione e delle connesse procedure di formalizzazione.

Almeno un cenno desidererei assicurare al gran tema del modellamento della voce e della scrittura, delle rispettive acquisizioni / apprendimenti, della loro tematizzazione metalinguistica e della conseguente trasposizione sul piano tecnologico.

Nella mostra cui accennavo eravamo partiti dalle articolazioni della voce registrate sugli oscillogrammi per mostrare poi – grazie a tecniche ostensive ben più antiche – scritture diverse, alfabetiche e ideografiche¹⁴, notazioni matematiche, logiche, informatiche¹⁵.

Mostrare, allineare, giustapporre gli oggetti raccolti era possibile per la loro natura di entità appartenenti al mondo fisico: dalle emissioni di voce articolata dall'apparato fonatorio umano all'argilla delle tavolette segnate dallo stilo, dal papiro ai gusci di tartaruga, al legno cerato incisi, dalla pergamena alla carta vergate con l'inchiostro dalla mano umana, dai caratteri metallici mobili per la stampa ai cristalli liquidi.

È stato possibile così vedere imprimersi nel fluire della voce, nello spazio di una tavoletta, della pelle di un animale, in una foglia di papiro, forme la cui fruizione è impensabile senza una chiave di accesso, di decifrazione o di lettura: indispensabile per comprenderle, in quanto non mere modulazioni della capacità articolatoria della voce, non pure evoluzioni dello stilo o semplici macchie d'inchiostro, ma veri e propri messaggi con un senso assegnato, entità storiche, culturali, opere uma-

¹⁴ E. Banfi, M.-D. Popelard, *Peindre les idées. Sur la calligraphie chinoise*, Presses Universitaires de France, Paris 2007.

¹⁵ S. Crespi Reghizzi, *Formal Languages and Compilation*, Springer, New York 2009.

ne consegnate allo scambio comune. A determinare ciò che trasforma un fascio di onde sonore in un messaggio vocale, una superficie incisa, impressa o configurata in un messaggio scritto puntano da gran tempo le scienze del linguaggio.

5. ANALOGICO E DIGITALE

Proprio in merito alle ragioni di tali evoluzioni insieme al permanere di finalità costanti e fondamentali, già nel 2000 in uno dei Seminari di Scienze del Linguaggio promosso dal nostro Dipartimento di Scienze Linguistiche, Luca Mari, collega di *Fondamenti di informatica*, era intervenuto sul significato di *analogico* e *digitale*. Il tema è stato poi ripreso e sviluppato: nel 2002 con la pubblicazione di un libro, *Atomi & bit. Le ragioni del digitale e del multimediale*¹⁶, nel 2004 con una comunicazione a più voci all'VIII Congresso internazionale di Semiotica a Lione¹⁷, cui a breve seguirà un articolo¹⁸ su una rivista internazionale.

La questione è, nella sua integralità, semiotica.

Proverò a delineare in massima sintesi la tesi centrale: non è sufficiente la condizione sintattica per l'esistenza dell'informazione, formalizzata a suo tempo da Hartley e Shannon, in base alla quale perché ci sia informazione occorre che l'entità che porta informazione sia scelta entro un insieme contenente almeno due entità. Condizione, questa, direttamente replicata a proposito dei requisiti della controparte fisica: un sistema fisico può essere adottato come supporto per l'informazione se può assumere almeno due stati riconoscibili come distinti all'osservazione. Ma tale condizione non è sufficiente perché molto spesso le informazioni da processare implicano il riconoscimento non solo di dif-

¹⁶ L. Mari, *Atomi & bit. Le ragioni del digitale e del multimediale*, Guerini e Associati, Milano 2002.

¹⁷ A. Frigerio, L. Mari, M. Padula, S. Raynaud, *The Analogue/Digital Opposition: A Semiotic Matter*, 8^e Congrès de l'Association Internationale de Sémiotique (AIS), Lyon 2004.

¹⁸ A. Frigerio, A. Giordani, L. Mari, *On representing information: a new characterization of the analogue / digital distinction*, (i.c.s.) 2010.

ferenze, ma anche di relazioni più complesse, relative non solo alla cardinalità ma anche all'ordinalità degli elementi strutturati. Ora, quando un insieme di informazioni è provvisto di struttura, dunque le entità che lo costituiscono veicolano sia informazione sia meta-informazione (cioè informazione sull'informazione), si pone un problema generale: come mantenere tale meta-informazione nell'operazione di scrittura, di codifica, così da essere in grado di leggerla (decodificarla) correttamente dal sistema fisico adottato come supporto? La risposta si può riassumere così: *analogica* è la strategia grazie alla quale la meta-informazione è *trasferita al supporto*, cosicché sia la scrittura che la lettura corrispondono all'applicazione di una mappatura omomorfa (che cioè preserva la struttura); *digitale* è invece la strategia grazie alla quale la meta-informazione è fissata nella regola di scrittura / lettura, così che al supporto fisico si richiede soltanto di essere in grado di assumere almeno due configurazioni distinguibili (usualmente indicate come 0 e 1). Basti pensare, a titolo d'esempio, alla differenza tra orologi analogici e digitali. Si tratta dunque della possibilità di separare il *medium* dal messaggio, proprio del digitale.

6. LINEARITÀ APPARENTI, COMPLESSITÀ EFFETTIVE

Dunque è determinante la competenza del lettore, nell'atto di lettura. Quanto più complesso è il testo da leggere, tanto più alta la competenza. In un'epoca come la nostra, caratterizzata dalla cosiddetta svolta cognitiva in molti ambiti disciplinari, si può osservare l'incidenza – sulla comprensione e valutazione di pratiche culturali antiche e caratterizzanti l'umano, come il leggere e lo scrivere – di sensibili sviluppi delle conoscenze genetiche e neurologiche relative al corpo, al cervello, alle relative attività.

Ho seguito recentemente con grande interesse la presentazione del volume di Marianne Wolf, *Proust e il calamaro. Storia e scienza del cervello che legge*. Non siamo nati per leggere, sostiene l'autrice, neuroscienziata che dirige, a Boston, il *Center for Reading and Language Research*. Ma siamo dotati di un cervello straordinariamente plastico, ot-

timo esempio di architettura aperta che, per imparare a leggere, ha dovuto e ancora oggi deve creare sofisticati collegamenti fra strutture e circuiti neuronali in origine preposti ad altri e più basilari processi, come la vista e la lingua parlata. Il cervello, riplasmato dalla lettura, ha così consentito la formazione di un sapere caratterizzato dall'accumulo, creativo e vertiginosamente efficace, di sempre nuove conoscenze. Nell'attuale fase di transizione, caratterizzata per i *nativi digitali* da un più immediato accesso alla rete, come si riconfigurerà – si chiede la Wolf – il ruolo della lettura, di una lettura senza immagini, libreria? Vi si affezioneranno ancora? Al cuore stesso del processo di lettura sta, secondo l'autrice, la capacità di andare oltre il testo, pensando con la propria testa ai mondi dischiusi dalla lettura. Pensando a quei mondi, scrive la Wolf, si pensa a sé. È l'esperienza della cosiddetta *lettura profonda* o *lettura esperta*:

*L'interazione dinamica tra esperienze di lettura e di vita è bidirezionale: apportiamo al testo le nostre esperienze, ma le letture influiscono sulle nostre esperienze di vita. Pochi autori hanno colto la reciprocità di questa relazione meglio di Alberto Manguel in Una storia della lettura: l'intero libro è una ricostruzione di come lui e il testo si trasformano a vicenda. A volte emergiamo come Manguel da questa discesa nel mondo del pensiero con una più ampia capacità di pensare, sentire e agire in modi nuovi e arditi. Ma dovunque siamo stati condotti, non saremo più come prima. Esistono correlati fisiologici di questa esperienza che suggeriscono cambiamenti neurofisiologici quando il livello del lettore esperto è stato raggiunto.*¹⁹

Ma non è solo la lettura di Proust a rapirci. Ciascuno potrebbe indicare quali letture lo hanno segnato, quali testi gli abbiano dischiuso mondi prima sconosciuti. Dunque nell'accesso a un testo si può procedere per tappe e approfondimenti successivi.

¹⁹ M. Wolf, *Proust e il calamaro. Storia e scienza del cervello che legge*, Vita e Pensiero, Milano 2009, pp. 174-175.

7. UTENTI ED ESPLORATORI.

IL RUOLO PIONIERISTICO DI ROBERTO BUSA

Se da un lato al cambiare delle modalità di presentazione di messaggi e testi cambiano anche le modalità di fruizione, non è escluso che l'utente avvertito dell'impatto su di sé di tali modalità non voglia passare dalla condizione di utente-guidato a quella di esploratore sul campo. Anzi, c'è chi ha precorso i tempi e ha forzato la mano a chi non sognava ancora la possibilità di destinare macchine al lavoro dell'umanista, lasciandosi tuttavia coinvolgere.

È il caso di P. Roberto Busa, gesuita vicentino ormai novantaseienne (1913-), che la bellezza di sessant'anni fa (1949) otteneva dall'amministratore delegato dell'IBM, Thomas J. Watson (1874-1956), non senza vincerne qualche resistenza, sostegno a un progetto che si sarebbe rivelato pionieristico, tanto da rendere P. Busa *ricosciuto universalmente come il pioniere dell'informatica linguistica*²⁰. Padre Busa era già alle prese con il monumentale lavoro di lemmatizzazione di tutti gli scritti di san Tommaso, e degli autori a lui collegati, lavoro iniziato nel 1947, come esito della sua tesi di dottorato alla Gregoriana. Gli lascio volentieri la parola, quale si legge nell'Introduzione al volume che, prima pubblicazione sua, è anche l'esito della medesima ricerca di dottorato: *La terminologia tomistica dell'interiorità. Saggi di metodo per un'interpretazione della metafisica della presenza*:

L'indagine da me compiuta su san Tommaso, dei cui risultati si compone questo mio libro, e vari sondaggi praticati a ricercare in Aristotele, Plotino e sant'Agostino, Alessandro di Hales, sant'Alberto Magno e san Bonaventura, non tanto gli enunciati teoretici

²⁰ Voce *Roberto Busa* nell'*Enciclopedia Treccani*, appendice 2000, p. 260. Cfr. anche la prefazione di A. Bausola al volume di R. Busa *Fondamenti di informatica linguistica*, Vita e Pensiero, Milano 1987, pp. 9-10 e l'articolo di G. Bolognesi, *La linguistica computazionale nell'Università Cattolica del S. Cuore e l'origine del termine informatica*, in «Aevum» 73/3 (1999), pp. 914-920. Del riconoscimento *extra moenia* dell'operato scientifico di P. R. Busa può bastare, qui, ricordare che nel 1998 la *Association for Literary and Linguistic Computing* (ALLC) e la *Association for Computers in the Humanities* (ACH) hanno istituito il *Busa Prize* (<http://www.digitalhumanities.org/viewAdho/TheBusaPrize>) che premia chi si è distinto nel campo dell'informatica umanistica.

propri a ciascuno, quanto direttamente l'impiego dei mezzi espressivi scelti a significarla, porteranno a concludere [sottolineatura nostra] *che l'interiorità per la Scolastica e per San Tommaso spazia in tre piani, profondamente diversi, pure, per ciò stesso, inscindibilmente collegati: l'interiorità ontologica o del predicato nell'essere o dell'idea nell'idea; l'interiorità spirituale dell'anima che ricerca Dio nel proprio cuore.*

Il dispiegamento dei segni verbali che regge la complessità di questa triplice trama risulterà veramente imponente [evidenziazione nostra]. *Vi si distinguono quattro solchi principali: 1) la terminologia propria e formale dell'interiorità: in con i suoi composti, opposti, sinonimi e affini, anzitutto in quanto esprime i rapporti spaziali dei corpi tra loro e con il luogo; secondariamente in quanto viene esteso ai rapporti tra enti in quanto tali e perciò anche tra spiriti; 2) la grossa famiglia di voci che esprimono l'attività e la partecipazione; 3) le voci che esprimono immagine, imitazione, somiglianza, differenza, uguaglianza, diversità, ecc. ...; 4) le voci riferentisi alla luce, con le sue qualità e funzioni, nel loro aspetto fisico-naturale e nel loro simbolismo mistico-spirituale.*²¹

Come procedere?

Il programma da svolgere – prosegue Busa – mi si presentò, come è ovvio, distinto in due momenti: considerare anzitutto il testo assolutamente, secondo quanto esprime da sé in sé; di poi studiarlo nelle relazioni della sua antecedente e conseguente ambientazione storica. Solo così il pensiero dell'Angelico verrebbe esaurientemente esposto. Ma dopo alcuni assaggi nell'uno e nell'altro senso, sgomentato dalla sua vastità – vi si incontrano infatti le principali ascese del pensiero: Dio, mondo, anima – ebbi chiara coscienza che, se avessi voluto affrontare ambedue quelle parti, le mie conclusioni in ambedue sarebbero state affrettate. Rinunciai quindi al secondo lavoro ... risolutamente volli evitare che il mio fosse, come rimproverava Kant nell'Introduzione ai Prolegomeni a ogni futura metafisica, solo un progetto di lavoro scientifico. Piuttosto che sfiorarne a

²¹ R. Busa S.J., *La terminologia tomistica dell'interiorità. Saggi di metodo per un'interpretazione della metafisica della presenza*, Bocca, Milano 1949, p. 6.

pizzico le singole pari, ebbi l'ambizione di svolgerne non fosse che una sola, non importa se minima, ma, per quanto lo potessi, esaurivamente ...

Ciò posto, restava a vedere quale metodo ne avrebbe garantito la verità ...

Ora appare evidente che radice di gran parte delle difficoltà, oltre all'arduità propria di ogni dottrina filosofica, è il non dominare sufficientemente la terminologia dell'autore. Non mi riferisco a quelle parole, come uomo, pianta, carro, il cui significato, indicando esse la globalità d'una complessa, solida sostanza materiale, oggetto d'una comune intuitiva cognizione, è sempre sufficientemente individuabile, ma alle parole che esprimono realtà più semplici, le quali, elementi costitutivi e vari aspetti delle precedenti, vengono direttamente considerate solo in una conoscenza riflessa, come interpretazione di quelle, ad es. idea, forma, moto, qualità, abitudine e sono oggetto proprio della filosofia ...

Sono tali parole che racchiudono l'ossatura di una filosofia ... Per questo nel titolo del mio lavoro ho parlato d'interpretazione, perché ritengo che ogni vera filosofia debba andare interpretata ...

Ma come allora leggere in un testo ciò che le parole immediatamente non sanno dire? ... Giudicai che fosse non soltanto necessario a superare la zona delle interpretazioni approssimative e passibili di impugnazione, ma anche sufficiente a darci quella assimilazione dei principi vitali, senza la quale non si arriva a una sistemazione definitiva di un corpo di dottrina, il pieno adeguarsi alla terminologia dell'autore, il riprodurre in sé, il più integralmente possibile, sia quanto a numero sia quanto a mutui rapporti e proporzione, tutto il complesso dei segni espressivi, dei quali si vestiva quell'interiore mondo spirituale di concetti ...

Scindere pezzo a pezzo gli elementi del parlare di S. Tommaso e chiederci di ciascuno: che significa per lui questa parola? Ecco quale fu in sostanza nel suo primo passo il mio metodo.

Il mio caso portava che mi occupassi di quel gruppo di voci che esprimono formalmente il rapporto d'interiorità. Mi proposi perciò di

seguire una a una ciascuna d'esse, scorrendo riga per riga il testo, considerando: 1) tutte le volte in cui apparisse, fosse pure ad altro proposito; 2) le parole da essa derivate, a essa connesse o affini e le voci opposte; 3) il suo alternarsi e collegarsi, nella sua proporzione e frequenza con le altre.

Il lavoro quindi consisteva anzitutto nella schedatura dei testi riportanti quella data espressione. Veniva poi il loro smistamento in base ai vari significati che la parola assumeva, per arrivare finalmente a una catalogazione sistematica di tutti e a un'analisi di ciascuno ...

*Il lavoro così s'è fatto denso, duro e pesante: e non poteva non esserlo. La prima parte si dilunga anche in punti non attinenti direttamente al nostro argomento: ma essendo il mio impegno quello di seguire in tutti i loro usi le singole voci, mi dovetti in ciò lasciar guidare dal testo: era bene pagare a questo prezzo la garanzia che tutto il materiale interessante lo scopo venisse interamente raccolto e analizzato.*²²

Ho preferito indugiare sulla genesi di un progetto a cui il temuto rimprovero kantiano non può essere mosso: non si sarebbe trattato solo di un progetto di lavoro scientifico, ma di un lavoro realizzato: con la pubblicazione dell'*Index Thomisticus*²³ in 56 volumi, nel 1980, con la sua trasposizione su CD nel 1989; infine, dal 2006, nella versione consultabile on-line all'indirizzo: <http://www.corpusthomisticum.org/>.

Da quella realizzazione sarebbero scaturite metodologie che hanno potuto essere estese a molti altri progetti in lingue e alfabeti diversi, progetti che si sono potuti avvalere di tecnologie in continuo sviluppo, su supporti via via diversi e sempre migliori: dalle schede perforate ai nastri magnetici ai CD ai DVD alla rete. Compiuta questa prima fase, per un certo tempo P. Busa ha annunciato che un'altra fase di lavoro si delineava all'orizzonte, quella del compimento lessicografico dell'opera, capace di corredare i lemmi dell'*Index* dei rispettivi significati, tenendo

²² Ivi, pp. 6-10.

²³ R. Busa S.J., *Index Thomisticus Sancti Thomae Aquinatis Operum Omnium Indices et Concordantiae*, Frommann Holzboog, Stuttgart 1974-80, 56 volumi.

conto dei cambiamenti culturali intervenuti: il progetto chiamato *Lessico Tomistico Biculturale* ²⁴.

8. LINGUISTICA COMPUTAZIONALE E PROGETTI IN SVILUPPO

Nel 2006 ha preso avvio la realizzazione di una nuova fase di ricerca, condotta all'Università Cattolica di Milano, presso il *Centro di Ricerche Interdisciplinari per la Computerizzazione dei Segni dell'Espressione* (C.I.R. C.S.E.) e coordinata da Marco Passarotti. Si tratta di una ricerca volta a porre le basi metodologico-formali, sul piano linguistico e informatico, teorico e applicativo, per il trattamento automatico dei contesti dei lemmi dell'*Index*. Come primo lemma da trattare è stato scelto *forma*; è stato individuato nel formalismo delle *Dependency Grammars* l'ambito teorico di riferimento per la descrizione sintattica di una lingua altamente flessiva e di natura non-proiettiva quale il latino; nei metodi sviluppati per la costituzione del *Prague Dependency Treebank* (PDT) si è riconosciuto un alto grado di applicabilità alle strutture della lingua latina. È dunque in corso l'annotazione sintattica manuale e in modalità *computer-assisted* delle migliaia di frasi (15.000 occorrenze) in cui occorre in Tommaso un termine carico di storia filosofica come *forma*. I dati e la loro elaborazione sono consultabili all'indirizzo <http://itreebank.marginalia.it/>.

A prescindere dalle molte precisazioni tecniche che si potrebbero apportare, ma non sono compatibili con il taglio e l'estensione di questo articolo, vorrei soffermarmi su un paio di punti. Voler affondare lo sguardo nell'opera di un autore ha imposto e sta rendendo necessario un trattamento sistematico di tutti i livelli linguistici di quei testi. In particolare, ha reso e rende necessario passare attraverso la morfosintassi per procedere dal censimento di tutte le voci lessicali all'elaborazione dei rispettivi valori semantici.

²⁴ M. Passarotti, *Verso il Lessico Tomistico Biculturale*, in R. Petrilli, D. Femia (a cura), *Il filo del discorso. Intrecci testuali, articolazioni linguistiche, composizioni logiche*, Atti del XIII congresso nazionale Viterbo 2006, Pubblicazioni della Società di Filosofia del Linguaggio 04, Aracne, Roma 2007, pp. 187-205.

Nel frattempo, dagli albori dell'intuizione – ancora negli anni Quaranta del Novecento – a oggi, è maturata tutta una comunità scientifica, quella della linguistica computazionale²⁵, che ha preso a occuparsi delle più varie lingue del mondo. Basti qui accennare almeno ad alcune delle grandi strutture o dei grandi appuntamenti che attestano la vitalità di questo filone di ricerca: l'*Association for Computational Linguistics* (ACL: <http://www.aclweb.org/>), che nel 2012 terrà il suo 50° incontro annuale, le *International Conferences on Computational Linguistics* (COLING: <http://www.coling-2010.org/>), nel 2010 al 23° appuntamento, le *Conferences on Language Resources and Evaluation for Human Language Technologies* (LREC: <http://www.lrec-conf.org/lrec2010/>) iniziate nel 1998, a ritmo biennale.

Lo scenario che si è venuto delineando è globale, anche se non immune dal *digital divide*. E l'affinamento dei processi di elaborazione delle informazioni, insieme alla crescente disponibilità di dati, sta orientando la ricerca anche in direzione dell'aspetto più *contenutistico* del fatto linguistico, verso la semantica.

9. SCIENZA, FILOSOFIA E L'ARTE DELL'INTERPRETAZIONE

Eccoci dunque ricondotti a considerare gli obiettivi originari, alla base del ricorso al computer nel trattamento delle lingue e dei testi, e a confrontarli con quelli attualmente perseguiti. La lettura in profondità, sistematica e totale delle fonti non sembra più, a sessant'anni dalla formulazione di un primo programma, soltanto un sogno. È nata e sta crescendo una giovane scienza, che non sopprime tuttavia – credo di poter dire – l'esigenza di quell'arte sapiente che è richiesta dall'interpretazione testuale.

Perché arte, perché arte sapiente? Perché – secondo chi scrive – i testi, che sono intessuti non solo di parole, ma anche di pensieri, sono intrisi di mondo: sono l'esito di atti di enunciazione che li radicano nel

²⁵ A. Lenci, S. Montemagni, V. Pirrelli, *Testo e computer. Elementi di linguistica computazionale*, Carocci, Roma 2005.

tempo e nello spazio, da cui pure possono avere preso congedo da gran tempo per continuare la loro opera, circolando in tempi e spazi altri. Coinvolgono dunque culture e storie ora prossime ora remote e riaprono l'orizzonte dell'indagine, oltre quell'autotelismo del segno testuale e linguistico di cui dicevamo all'inizio, in direzione di *telo*i ulteriori, di quell'acquisizione, elaborazione e gestione di conoscenze che rende la semiosi umana un *medium* strategico in ordine al pensare e all'agire, personale e sociale.