

Scritture dimenticate, scritture colonizzate: sistemi grafici e codifiche digitali nelle culture araba e indiana

Paolo Monella

Introduzione

Le società che elaborano una tecnologia, la modellano sulla loro cultura. Le altre, subiscono la pressione a rimodellare la propria cultura su quella tecnologia. Questo breve saggio intende presentare alcune riflessioni su come le attuali tecnologie di codifica digitale del testo rappresentino e manipolino due sistemi grafici non-occidentali, quello indiano (devánāgarī) e quello arabo.¹

Per Raul Mordenti

Alle radici di questo studio c'è l'insegnamento di Raul Mordenti. Quasi quindici anni fa proprio i suoi scritti, insieme a quelli di Tito Orlandi, Dino Buzzetti, Giuseppe Gigliozzi, Domenico Fiormonte e degli altri esponenti della 'scuola romana' di informatica umanistica hanno costituito la mia prima introduzione alla disciplina. In quei saggi degli anni Novanta/Duemila, con la garbata ostinatezza di chi non teme di andare contro corrente, Raul Mordenti ci invitava, come continua ad invitarci oggi, ad uno studio critico-metodologico delle nostre rappresentazioni digitali dei sistemi semiotici e culturali.²

Un carattere *a* in un'edizione digitale non è, semplicemente, la *a* di un manoscritto latino medievale di cui quell'edizione tiene conto, o che trascrive.

Quel carattere è un segno all'interno di un ben preciso *charset* (o tabella di caratteri, ad esempio Unicode/*Latin script/Basic Latin*), inserito in un documento con una precisa codifica (ad esempio TEI XML P5). Tanto il *charset*

¹Ho accennato più brevemente alle idee sviluppate in questo saggio già in DOMENICO FIORMONTE, DESMOND SCHMIDT, PAOLO MONELLA E PAOLO SORDI, «The politics of code. How digital representations and languages shape culture» (par. *Universalizing the typography: Unicode*) e in PAOLO MONELLA, *Forme del testo digitale*, in Raul Mordenti (a c. di), *Filologia Digitale. Problemi e Prospettive*, Roma, Accademia Nazionale dei Lincei; Bardi Edizioni, 2017, vol. 135, pp. 148-150.

²Una sintesi delle riflessioni di quel periodo che penso sia utile rileggere anche oggi — per le sue dense riflessioni teoriche, che non invecchiano quanto le rassegne di strumenti e tecnologie recenti — è RAUL MORDENTI, *Informatica e critica dei testi*, Roma, Bulzoni, 2001.

quanto la codifica sono sistemi semiotici. In un'edizione critica digitale, poi, il segno-carattere digitale *rappresenta* un altro segno, ovvero un grafema (o un allografo, o un idiografo di uno scriba) del manoscritto. A sua volta, quel grafema è un elemento di un preciso sistema grafico medievale (un sistema semiotico complesso e autonomo, con le sue convenzioni abbreviative, i suoi diacritici, la sua punteggiatura) che a sua volta ha alle spalle almeno due diversi sistemi semiotici: un alfabeto e una lingua.³

Dunque il carattere digitale è un segno che rappresenta un altro segno. E nessuno dei processi semiotici che stanno dietro la nostra rappresentazione digitale del testo è neutro: ogni scelta (cosa e come rappresentare) ha aspetti teorico-metodologici che il filologo digitale deve tenere presenti, esplicitare e motivare se vuole fare ricerca.⁴

Finora nella mia ricerca ho cercato di applicare questo imperativo metodologico alla nostra rappresentazione dei sistemi grafici dei manoscritti medievali nell'ambito dell'edizione critica digitale.⁵ Insomma, i sistemi semiotici (grafici) 'altri'

³L'impostazione del presente saggio presuppone il principio fondativo della 'grafematica', ovvero che i sistemi grafici costituiscano sistemi semiotici autonomi, da studiare in sé oltre che nel loro rapporto con la fonetica e il sistema linguistico in generale. Si tratta di un principio dimostrato a mio parere in modo assai convincente da chi si è occupato specificamente di questo ambito (mi limiterò a citare solo JACQUES DERRIDA, *De la grammatologie*, Paris, Les editions de minuit, 1967), per quanto ancora non universalmente acquisito. Ma l'idea dell'indipendenza della 'sostanza grafica' da quella fonetica nel manifestare la 'forma linguistica' (senza passare attraverso un rispecchiamento della fonetica) è già anche in LOUIS HJELMSLEV, *Prolegomena to a theory of language*, Madison, University of Wisconsin Press, 1961, cap. 21. Cito dalle pp. 103-104: «Thus the same linguistic form may also be manifested in writing [...] Here is a graphic "substance" which is addressed exclusively to the eye and which need not be transposed into a phonetic "substance" in order to be grasped or understood».

⁴Tra gli studi più recenti in cui Raul Mordenti ha sviluppato queste tematiche vorrei ricordare almeno RAUL MORDENTI, *Paradosis. A proposito del testo informatico*, Roma, Accademia Nazionale dei Lincei, 2011, che contiene tra l'altro (a pp. 640-648) una delle riflessioni a mio parere più lucide sulla questione dell'alfabeto (solo apparentemente semplice) e RAUL MORDENTI, *La trascrizione nella procedura ecdotica-informatica*, «Testo e Senso», 15, 2014. Resta comunque fondamentale il suo lavoro metodologico in vista dell'edizione dello Zibaldone Laurenziano: vd. RAUL MORDENTI, *Nani sulle spalle di giganti. Alcune lezioni di metodo della Filologia italiana e le prospettive dell'edizione critica digitale dello Zibaldone Laurenziano di Giovanni Boccaccio*, in *Annuario delle attività 2012 - Contributi del Centro Linceo Interdisciplinare beniamino Segre - vol. n. 131*, Scienze e Lettere, 2012, pp. 367-410; RAUL MORDENTI, *Filologia digitale (a partire dal lavoro per l'edizione informatica dello Zibaldone Laurenziano di Boccaccio)*, «Humanist Studies & the Digital Age», 2-1, 2012, pp. 37-56 e RAUL MORDENTI, *Edizione Critica Iper-testuale dello Zibaldone Laurenziano (Pluteo XXIX.8) autografo di Giovanni Boccaccio*, <http://rmcisadu.let.uniroma1.it/boccaccio/>.

⁵Da questa linea di ricerca è nata la mia edizione critica digitale di Orso Beneventano: PAOLO MONELLA, *Ursus from Benevento, De nomine, from the Aabbreviatio artis grammaticae, codex Casanatensis 1086, ff. 1r-11r*, Palermo, ALIM - Archivio della Latinità Italiana del Medioevo; Website; ALIM - Archivio della Latinità Italiana del Medioevo, 2017. Le principali pubblicazioni in cui ho esposto i principi metodologici sperimentati nell'edizione di Orso, oltre alla documentazione su GitHub dell'edizione stessa, sono PAOLO MONELLA, *Livelli di rappresentazione del testo nell'edizione del De nomine di Orso Beneventano*, «Umanistica Digitale», 2, 2018, pp. 67-91 e PAOLO MONELLA, *A sustainable work flow for a multi-layer edition of the Chronicon by Romualdus Salernitanus*, in Daria Spampinato (a c. di), *AIUCD2018 - book of abstracts*, Bologna, AIUCD; Alma DL UniBo, 2018, pp. 21-24. Tra le migliori riflessioni sulla

venivano del passato, dagli scrittori latini del medioevo europeo. Per questo saggio, invece, stimolato da Domenico Fiormonte, mi sono provato ad applicare lo stesso metodo critico a scritture (e quindi culture) ‘altre’ non storicamente, ma geograficamente, quelle indiane e arabe.

Questo studio presenta dunque un’esplorazione preliminare di un ambito di ricerca potenzialmente amplissimo ma, mi pare, poco esplorato. Io stesso non so se potrò in futuro portarlo avanti molto oltre le considerazioni suggerite qui, non essendo uno specialista di lingue e scritture orientali. Ma mi pare che valga la pena di tentare l’impresa, per stimolare una maggiore consapevolezza del significato culturale delle nostre rappresentazioni digitali, sia nella ricerca informatico-umanistica, sia nell’applicazione ‘di massa’ delle tecnologie informatiche.

Tre principi gutenberghiani

L’avvento della tecnologia della stampa in Europa nell’età moderna ha lentamente rimodellato il sistema grafico latino, come quello greco e quello cirillico, su un modello rigidamente ‘alfabetico’, ovvero sulla base di tre principi che valevano alle origini degli alfabeti greco e latino, ma già non più per i complessi sistemi chirografici, ovvero di scrittura manoscritta, sviluppatasi in età medievale:

- **1** 1. Ad ogni grafema corrisponde una sola lettera alfabetica, e possibilmente un solo fonema;
- **1** = **1**. Ogni grafema alfabetico vale quanto un altro (ad es. vocali e consonanti, o vocali lunghe e vocali brevi);
- **1, 2, 3...** La scrittura si struttura come una sequenza unidimensionale e unidirezionale di elementi tutti sullo stesso ‘livello’ (sia come valore sia come altezza nella riga). C’è solo un ‘prima’ e un ‘dopo’, che riflettono la sequenza nella catena fonica; non c’è, ad esempio, un ‘sopra’ e un ‘sotto’, e non si torna mai indietro.

Principi non universali

Questi tre principi non valevano per i sistemi grafici dei manoscritti del medioevo europeo, e ancor oggi non valgono per molti sistemi grafici non occidentali, neanche nella loro versione a stampa. Per fare solo qualche esempio:

- **1** 1. Nei manoscritti latini medievali una sola abbreviazione corrisponde sistematicamente a più lettere alfabetiche: ad es. \bar{u} in $lup\bar{u}$ sta per le due

grafematica digitale si possono citare TITO ORLANDI, *Informatica testuale. Teoria e prassi*, Roma, Laterza, 2010, capp. I e II; FREDERIKE NEUBER, *Das Konzept einer ,typografiezentrierten‘ digitalen Edition der Werke Stefan Georges samt einem Modell zur Beschreibung von Mikrotypografie* tesi di dottorato, Universität Köln, 2019 e l’articolo in corso di stampa PETER A. STOKES, *Aligning archetype: Towards a formal model for a transversal palaeography*, in Isabelle Marthot-Santaniello (a c. di), *Neo-palaeography: Analysing ancient Greek and Coptic handwritings in the digital age*, che ho potuto vedere in bozze.

lettere *u* e *m*, e (che può essere considerato un unico grafema) in *sona* sta per le lettere *p*, *e*, *r*.

- **1 = 1.** I segni diacritici, parte integrante dei sistemi grafici non rigidamente ‘gutenberghiani’, hanno uno status diverso da quello dei grafemi-base a cui si associano. Nei manoscritti latini questo si applica a segni abbreviativi come il macron in *lupū* per *lupum*. In quelli greci, a segni come lo iota sottoscritto () o lo spirito aspro () che, perso il valore fonetico che avevano nell’età classica, conservano valore distintivo nella grafia. Nel sistema grafico arabo, le ḥarakāt sono i diacritici che, aggiunti sopra o sotto un grafema consonantico, indicano la vocale breve con cui esso va pronunciato: da un punto di vista occidentale, potremmo considerarli grafemi come gli altri, relativi alle vocali (*a*, *i*, *u*) brevi, ma nella concezione araba non hanno lo stesso status dei grafemi consonantici.
- **1, 2, 3...** In tutti gli esempi riportati fin qui, i segni ‘modificatori’ (abbreviativi e diacritici) vanno posti sopra o sotto un grafema-base con cui si combinano. Nel sistema grafico devānāgarī diffuso in India, similmente a quello arabo, le vocali hanno uno status distinto dalle consonanti, tant’è che si combinano ad esse in legature, creando segni composti che indicano intere sillabe. In tali segni sillabici, la vocale si può trovare sotto, a destra oppure a sinistra della consonante. Dato che la scrittura devānāgarī si sviluppa da sinistra a destra, come quella latina, se la vocale si trova a sinistra della consonante ci troviamo di fronte ad un caso in cui la unidirezionalità della scrittura, che ci appare ‘naturale’ a causa del nostro punto di vista occidentale, viene contraddetta da un apparente ‘ritorno indietro’ verso sinistra. In realtà non si tratta di una inversione di direzione, ma di una conseguenza del fatto che non tutti i segni hanno lo stesso status (non è vero che 1 = 1): i grafemi vocalici modificano quelli consonantici. Nei casi discussi in questo paragrafo, dunque, la scrittura non si sviluppa su una sola dimensione (l’asse orizzontale, avanti/indietro) ma su due (asse orizzontale e asse verticale, ovvero sopra/sotto), e non necessariamente in un’unica direzione.

Stampa e alfabeto: un influsso reciproco

Sia nell’Europa rinascimentale sia al momento dell’introduzione della stampa nell’Impero ottomano all’inizio del Settecento, i primi stampatori si sono inizialmente sforzati di riprodurre le specificità dei sistemi grafici manoscritti (abbreviazioni, legature, allografi, diacritici, collocazione dei segni su due dimensioni)⁶.

⁶Sul tema generale della ‘rivoluzione testuale’ introdotta prima dalla chirografia e poi dalla stampa e il suo impatto sulla concezione del testo si vedano almeno R. Mordenti, *Informatica e critica dei testi*, op. cit., pp. 129-132, che rimanda a sua volta a ARMANDO PETRUCCI, *Scritture della memoria e memorie dello scritto. Dall’ordine degli oggetti scritti al disordine della scrittura virtuale*, «Parole chiave», 9, 1995, pp. 83-92; e DOMENICO FIORMONTE, *Scrittura e filologia nell’era digitale*, Torino, Bollati Boringhieri, 2003 cap. 1 (specificamente sulla stampa vd. il par. 1.3). Sull’introduzione della stampa nell’Impero ottomano si veda J. R. OSBORN, *The type of calligraphy: Writing, print, and technologies of the arabic alphabet* tesi di dottorato, University of California San Diego, 2008, in particolare il cap. 4 per gli aspetti storici e il cap.

Ma la tecnologia della stampa gutenberghiana aveva una necessità intrinseca: utilizzare un insieme limitato di caratteri mobili combinati nella maniera più semplice, ovvero, laddove possibile, l'uno di seguito all'altro.

In Europa nei primi secoli dell'età moderna la tecnologia gutenberghiana ha progressivamente impresso la propria 'forma' a quell'aspetto specifico della cultura europea che era la scrittura alfabetica (manoscritta) greco-latina: abbreviazioni e legature sono andate scomparendo, imponendo ai sistemi grafici europei una più rigida aderenza a quei tre principi (1 1; 1, 2, 3...) che ormai ci sembrano così 'naturali'.

In fondo, però, su tali principi erano state fondate in origine, prima delle evoluzioni della chirografia, le scritture capitali greca e latina, progenitrici dei sistemi grafici europei moderni (latino-europeo occidentale, greco-bizantino, cirillico). Ad esempio, nelle lingue indoeuropee come il greco antico consonanti e vocali hanno pari dignità, cioè pari ruolo distintivo, dal punto di vista fonetico e morfologico: questa è una delle caratteristiche principali che hanno portato alla nascita della scrittura alfabetica in Grecia, a partire da quella sillabica di origine semitica (fenicia).⁷

L'influsso è stato dunque reciproco: da un lato la stampa ha semplificato le scritture occidentali, ma dall'altro anche la tecnologia della stampa come noi la conosciamo è stata determinata dal contesto della scrittura alfabetica, modellata a sua volta sulle lingue europee. In questo senso, si può dire che la cultura europea ha contribuito a modellare su di sé la tecnologia della stampa.

Da Gutenberg ad ASCII, attraverso l'inglese

Se l'applicazione della stampa ai sistemi grafici dell'India (devānāgarī) e del Medio Oriente (arabo) non ha avuto un enorme impatto 'normalizzante', cioè di fatto occidentalizzante, rischiano di fare di peggio le tecnologie digitali di codifica del testo che, sviluppate soprattutto negli Stati Uniti e in Occidente, sono basate sui principi della stampa occidentale.

L'esempio dei diacritici proposto sopra, nel secondo punto del paragrafo *Principi non universali*, se applicato ai sistemi grafici moderni mostra come i tre principi gutenberghiani siano incarnati principalmente dalla scrittura contemporanea dell'inglese, povera di diacritici. Accentti, dieresi e altri segni 'modificatori' dei grafemi-base sono invece strutturali in moltissimi altri sistemi grafici di origine greco-latina, usati per lingue che vanno dal francese al russo, dal greco al turco.

È però proprio intorno al sistema grafico dell'inglese che, nelle prime fasi dello sviluppo dell'informatica, la tecnologia statunitense ha costruito la codifica di caratteri American Standard Code for Information Interchange (ASCII), che ha costituito per decenni il principale standard digitale per l'interscambio di testo,

⁵ per una riflessione sul suo impatto culturale.

⁷Vd. G. SAMPSON, *Writing systems: A linguistic introduction*, Stanford, Stanford University Press, 1990, pp. 99-107 (pp. 100-101 in particolare).

e ancor oggi costituisce l'unico standard 'sicuro' perché intorno ad esso sono costruite le principali infrastrutture del web.⁸

Input, codifica, visualizzazione, processamento

Va premesso che mentre nella stampa, come nella scrittura a mano, il segno grafico viene codificato, conservato e decodificato solo tramite il suo aspetto visivo, nelle tecnologie digitali di trattamento del testo vanno distinti vari aspetti, tra cui:

- l'*input*, una serie di pratiche tramite cui lo scrivente umano comunica al computer quale 'carattere' vuole codificare, e lo fa normamente tramite l'ausilio di segni grafici su una tastiera reale o virtuale;
- la *codifica* vera e propria, invisibile allo scrivente, tramite cui il computer trasforma i caratteri indicati dallo scrivente in un numero identificativo, all'interno di un insieme determinato di grafemi (o *charset*, come ad esempio ASCII o Unicode);
- la *visualizzazione*, ovvero la ri-generazione di un segno grafico su una superficie bidimensionale (come uno schermo o un foglio stampato) a partire dal numero che costituiva la rappresentazione digitale interna al computer.

A questi va aggiunto un ulteriore aspetto, spesso trascurato:

- il *processamento* digitale dei numeri in cui i 'caratteri' sono stati codificati, che prescinde da input e visualizzazione: ad esempio, la ricerca di una stringa (una sequenza di caratteri, come *Beatrice*) all'interno di una base di dati testuale (ad esempio il testo digitalizzato della Divina Commedia, un database di fornitori, una raccolta di leggi o semplicemente una pagina web).

Le legature devānāgarī

Come abbiamo visto nel paragrafo *Principi non universali*, nella scrittura indiana devānāgarī le vocali si legano alle consonanti formando di fatto con esse un simbolo sillabico unico. In alcuni casi una vocale si lega, e quindi si scrive, non a destra ('dopo'), ma a sinistra ('prima') della consonante che modifica, contraddicendo il principio gutenberghiano della sequenzialità unidirezionale dei grafemi (1, 2, 3...). Ho sostenuto che non si tratti di una inversione arbitraria della direzione della scrittura, ma di una conseguenza del fatto che le vocali in questo sistema scrittorio hanno lo status di 'modificatori' delle consonanti (incompatibilmente col principio gutenberghiano $1 = 1$).

⁸Vd. DOMENICO FIORMONTE, *Per un critica culturale delle Digital Humanities*, «Quaderni DigiLab», 2-1, 2012, pp. 220-242 (soprattutto p. 224); DOMENICO FIORMONTE, *Towards a cultural critique of the digital humanities*, in Manfred Thaller (a c. di), *Controversies around the digital humanities*, Cologne, Historical Social Research/Historische Sozialforschung, 2012, vol. 37, par. 3 e DOMENICO FIORMONTE, TERESA NUMERICO E FRANCESCA TOMASI, *The digital humanist. A critical inquiry*, Punctum Books, 2015, p. 72.

Tra l'altro, la concezione indiana della fonetica che sta dietro a questo uso grafico appare pienamente razionale, non meno di quella retrostante agli alfabeti occidentali, se si consideri come nell'articolazione fonica consonanti e vocali abbiano effettivamente status molto diversi, e si articolino in strettissima correlazione reciproca nell'unità fonica della sillaba: «registrando la propria voce in un computer e utilizzando un software per il trattamento elettronico dello spettro vocale [...] non ci è possibile udire distintamente qualcosa che assomigli a /b/ o ad /a/ in isolamento: o sentiamo /ba/ oppure uno strano riverbero elettronico. Quindi, a livello fonetico, una sillaba non è scomponibile, lo è solo a livello di astrazione».⁹

Il modello di codifica del testo di ASCII o di Unicode, invece, ignora la distinzione di status e il rapporto reciproco tra vocali e consonanti nella scrittura devānāgarī, ma attribuisce a tutte indifferentemente un numero identificativo (*code point*) e le giustappone semplicemente le une alle altre. Una parola hindi codificata tramite le attuali tecnologie digitali viene dunque convertita in una sequenza unidirezionale di numeri/codici, tutti sullo stesso livello.

La contraddizione tra il modello eurocentrico retrostante e la specificità culturale indiana esplose quando si rappresenta una sillaba come *pi*. La sua codifica Unicode è la sequenza U+092A (*p*) U+093F (*i*), in cui la *i* viene dopo; ma in scrittura devānāgarī la vocale (*i*) va scritta a sinistra ('prima') della consonante *p*: la sillaba risultante sarà dunque *pi*. La 'inversione' viene effettuata dal software solo a livello di visualizzazione, nascondendo il fatto che la codifica impone un modello occidentale al sistema grafico indiano.

Traggo l'esempio delle legature indiane da un saggio illuminante di Antonio Perri,¹⁰ che discute anche altri casi di attrito tra l'alfabeto-centrismo di Unicode e le logiche interne di scritture non occidentali, ovvero le scritture da destra a sinistra, la 'doppia codifica' dell'hangul, i caratteri cinesi: «La mia tesi, in breve, è che la digitalizzazione della scrittura proposta da Unicode si è rivelata

⁹ ANTONIO BARONI, *La grafematica: teorie, problemi e applicazioni* Tesi di laurea specialistica, Università di Padova, 2009, p. 85 (vd. anche pp. 77-98 e 189-202). Si vedano inoltre D. Fiormonte, *Scrittura e filologia nell'era digitale*, op. cit., p. 32 (che riporta la celebre opinione di Hegel per cui «La scrittura alfabetica è in sé e per sé la più intelligente»), e par. 1.2; GIORGIO RAIMONDO CARDONA, *Antropologia della scrittura*, Torino, Loescher, 1981, pp. 21-22 e 33-35. Una revisione radicale del concetto stesso di fonema e della sua presunta superiore razionalità, alla base del «paradigma segmentale», è in FEDERICO ALBANO LEONI, *Dei suoni e dei sensi: Il volto fonico delle parole*, Bologna, Il mulino, 2009, cap. 4 e FEDERICO ALBANO LEONI, *Il fonema: realtà o illusione?*, in Mauro Falcone e Andrea Paoloni (a c. di), *La voce nelle applicazioni. Atti dell'VIII Convegno AISV 2012*, Roma, Dipartimento di Linguistica Università Roma Tre; Bulzoni, 2012, pp. 3-16.

¹⁰ ANTONIO PERRI, *Al di là della tecnologia, la scrittura. Il caso Unicode*, «Annali dell'Università degli Studi Suor Orsola Benincasa», 2, 2009, pp. 725-748. L'esempio delle legature devānāgarī è alle pp. 754-737, ed è stato ripreso da D. Fiormonte, «Per un critica culturale delle Digital Humanities», art. cit., pp. 232-233. Sempre sulle legature indiane, viste da un punto di vista grafematico (non informatico-umanistico), si veda anche A. Baroni, *La grafematica: teorie, problemi e applicazioni*, op. cit., p. 91.

come una sorta di iper-tipografia, poiché ha adottato e spinto sino alle estreme conseguenze la “logica dell’alfabeto” e la sua “tirannia”». ¹¹

Le ḥarakāt arabe

All’interno del sistema grafico arabo, ricco di segni diacritici, discuterò solo l’esempio delle ḥarakāt, i segni che, aggiunti sopra o sotto un grafema consonantico, indicano la vocale breve con cui esso va pronunciato.

Assumendo un punto di vista interno alla cultura araba, le ḥarakāt contraddicono tutti e tre i principi enunciati sopra:

- **1** **1.** Le ḥarakāt possono essere indicate oppure no. La norma è che non lo siano. In quest’ultimo caso, non è più vero che ad ogni fonema (vocale o consonante) corrisponde un grafema, e quindi un carattere (un *code point*, un numero) Unicode. La stessa parola, nella cultura araba, può essere scritta solo con consonanti e vocali lunghe, o anche (in casi particolari) aggiungendo le ḥarakāt alle consonanti.
- **1 = 1.** Le ḥarakāt hanno uno status inferiore a quello dei grafemi indicanti consonanti e vocali lunghe. Questi ultimi, infatti, costituiscono lo ‘scheletro’ della parola, percepito come sufficiente a identificarla. ¹²
- **1, 2, 3...** La scrittura araba, a mano o stampata, non si sviluppa affatto su una sola dimensione, ma i vari tipi di diacritici (i’jām e tashkīl, di cui fanno parte le citate ḥarakāt) vanno posti sopra o sotto un grafema-base con cui si combinano.

Unicode, invece, attribuisce alle ḥarakāt un *code point* alla pari dei grafemi consonantici e delle vocali lunghe: ad esempio, la consonante *beh* è identificata con U+0628 (cioè il numero esadecimale 0628, corrispondente al decimale 1576); la ḥarakah *fatha* , che indica una *a* breve, con U+064E. Anche le parole arabe, rappresentate tramite le attuali (occidentali) tecnologie digitali, vengono dunque ridotte ad una sequenza unidirezionale di numeri equivalenti.

¹¹A. Perri, «Al di là della tecnologia, la scrittura. Il caso Unicode», art. cit., p. 727. Vd. R. HARRIS, *La tirannia dell’alfabeto*, Viterbo, Stampa Alternativa & Graffiti, 2003. L’ultima versione dello standard Unicode è descritta in THE UNICODE CONSORTIUM, *Unicode 13.0*, The Unicode Consortium, 2020, <http://www.unicode.org/versions/Unicode13.0.0/>. Un’utile esposizione di alcuni suoi limiti, per quanto condotta con un approccio per lo più tecnico, è in PETER CONSTABLE, *Understanding unicode™. A general introduction to the unicode standard*, in *NRSI: Computers & writing*, SIL, 2001, par. 6 (per la questione delle legature consonante-vocale vd. par. 6.2 e n. 4; par. 8 e nota 19).

¹²G. Sampson, *Writing systems*, op. cit., p. 85: «The most important fact about the vowels of Hebrew and other Semitic languages, and the chief reason why it is less useful to mark them in writing than it is in Indo-European languages, is that to a large extent the linguistic contrasts realized by vowels are grammatical rather than lexical. This means that even if the contrasts are not recorded in writing they can to a very large extent be determined from context, and it also means that they tend to be less crucial for practical purposes of communication»; e p. 86: «Given the limited role of vowels as distinctive elements [...], a script which indicates only consonants is not unreasonably ambiguous in practice» nelle lingue semitiche. Vd. anche G. R. Cardona, *Antropologia della scrittura*, op. cit., pp. 44-45.

Attriti

Anche nel caso dell'arabo, come del devānāgarī, la forzatura non è apparente nella fase dell'input e in quella della visualizzazione, grazie a software che gestiscono la digitazione da tastiera (reale o virtuale) e la ri-composizione su schermo dei grafemi, legature e diacritici inclusi.

Esiste però un attrito tra il modello occidentale alla base della tecnologia e la cultura non occidentale cui il modello viene imposto, e questo emerge quando si passa dalla codifica all'elaborazione digitale — anche semplice — del testo.

Per rendere l'idea, farò riferimento all'esperienza dei parlanti di lingue europee, nei cui sistemi scrittorî pure esistono diacritici (accenti, umlaut etc.) che si possono combinare con altri caratteri: in italiano, francese o spagnolo, ad esempio, la *e* (Unicode U+0065) si può combinare con un accento acuto (Unicode U+0301). Vocale e diacritico possono codificarsi in Unicode tramite la sequenza dei due codici (sarà in visualizzazione che i due segni saranno riuniti), o con un *combining character* unico (vocale più diacritico, *é*) identificato da un terzo *code point* Unicode (U+00E9).

Ora, una delle elaborazioni più semplici del testo è la ricerca testuale semplice (*string matching*). Se un utente cercasse la stringa *tre* in una pagina web in lingua spagnola, con la funzione 'trova' del browser, troverebbe le istanze della parola *tren* ma non quelle della parola *trémula*, in quanto *e* e *é* sono identificati da *code point* Unicode diversi.

Allo stesso modo, un utente arabo che eseguisse una ricerca in una pagina web di una parola, digitandola con le ḥarakāt, avrebbe molte probabilità (a seconda delle funzionalità del software che sta usando) di non trovare le istanze di quella parola scritte senza le ḥarakāt — e viceversa. Il problema, originandosi dalla codifica, si riverbera su ogni forma di elaborazione del testo, dall'interrogazione di un database alle ricerche sul web, a meno che software specifici non siano stati istruiti a superarlo, moltiplicandosi il suo impatto con l'aumentare della complessità della tecnologia.¹³

Ne deriva un forte impatto 'normalizzante' nella direzione dell'uso dell'alfabeto latino (in realtà americano): solo di recente, e gradualmente, si vanno superando i problemi di incompatibilità tra i caratteri Unicode non latini, o più precisamente non-ASCII (American Standard Code for Information Interchange) con protocolli basilari come la codifica degli indirizzi internet e email, delle password,

¹³Cfr S. BHATTACHARYYA, *Words in a world of scaling-up: Epistemic normativity and text as data*, «Sanglap: Journal of Literary and Cultural Inquiry», 4-1, 2017, pp. 31-42: «We draw attention to an interesting problem that arises when the queried word in a tool like the Hathitrust Bookworm is one that is from a non-European language, but which occurs within European-language texts — with the word occurring in the text in the roman alphabet, that is, in transliterated form. We found that the occurrence of such words was being underreported, and sometimes not being reported at all». Qui la questione specifica è quella della traslitterazione in caratteri latini, che non coglie in modo efficace — neanche computazionalmente, oltre che culturalmente — l'identità della parola.

degli stessi nomi dei file. Ad esempio, chi frequenta le tecnologie digitali da tempi meno ‘Unicode-friendly’ ha acquisito come ‘buona pratica’ ormai permanente, diciamo così, per prudenza, quella di evitare in questi àmbiti (indirizzi internet etc.) i caratteri non-ASCII pure appartenenti a lingue europee (come le lettere accentate, gli umlaut, i caratteri greci).¹⁴

Arabīzī

Per quanto riguarda l’arabo, ho scritto che, in fase di input, tastiere fisiche e virtuali su schermi *touch* permettono di mascherare il problema della codifica ‘occidentalizzante’. Questo è vero quando lo scrivente ha davanti la tastiera di un computer e scrive in fuṣḥā, cioè nella variante ‘alta’ e unitaria della lingua che, integrata in età contemporanea con numerosi neologismi, discende direttamente dalla lingua del Corano.

Ma quando i giovani si scambiano messaggi in lingua vernacolare (cioè nelle varianti locali dell’arabo) scrivendo sullo schermo di un cellulare o di un tablet, tendono sempre più a scrivere l’arabo in caratteri latini, più semplici da inserire in quei contesti, in quanto i sistemi operativi di quei dispositivi sono stati disegnati in occidente, e successivamente adattati all’input in caratteri arabi.

Il codice risultante è comunemente chiamato *arabīzī* (arabi-anglizi, cioè arabo-inglese) o *Franco-Arabic*, e mescola lettere latine e numeri — come il 3, usato per la lettera ‘ayn’ che non ha corrispettivi né fonetici né grafici nell’alfabeto latino. In esso, le vocali brevi possono venire indicate (con *a*, *i*, *u*), e non viene usato alcun diacritico, ad esempio per distinguere le vocali brevi dalle lunghe. In pratica, la codifica usata è la più semplice, la più ‘standard’ e quindi la più restrittiva possibile, quella ASCII.¹⁵

Questo è un aspetto che spesso si sottovaluta: le tecnologie possono aiutare nella codifica e la trascodifica da/in sistemi semiotici diversi da quelli anglosassoni

¹⁴Vd. sopra il paragrafo *Da Gutenberg ad ASCII, attraverso l’inglese*. L’Internet Engineering Task Force (IETF) ha definito solo tra il 2008 e il 2010 il protocollo Internationalized Domain Names for Applications (IDNA) per rendere possibile la creazione di indirizzi web (‘domain names’) internazionalizzati (IDN), ovvero non composti solo da caratteri ASCII. Eppure ancor oggi, dieci anni dopo, indirizzando il proprio browser sul sito delle ferrovie austriache (ÖBB) tramite l’IDN <https://www.öbb.at/> (che include il carattere Unicode non-ASCII ö), si viene semplicemente reindirizzati all’URL ASCII <https://www.oebb.at/> (lo stesso per <https://www.süddeutsche.de/>, che rimanda a <https://www.sueddeutsche.de/>). Difficilmente un caporedattore di *Libération* tollererebbe l’errore di un giovane stagista che non segnasse l’accento sulla é, ma l’editorialista più raffinato della testata digita quotidianamente senza battere ciglio l’URL ASCII <https://www.liberation.fr/>, perché l’IDN <https://www.libération.fr/> semplicemente non esiste. Similmente, dal sito ufficiale del comune di La Coruña traspare grande attenzione per l’identità culturale galiziana: la sua lingua principale è il gallego (non il castigliano), e il suo indirizzo termina in .gal (non in .es). Se il sito resiste al centralismo spagnolo, soccombe però alla ‘tirannia dell’alfabeto’ americano, in quanto l’IDN <https://www.coruna.gal/> non esiste: i galiziani devono rassegnarsi ad un URL ASCII <https://www.coruna.gal/>.

¹⁵Cioè, per la precisione, il sottoinsieme ASCII di Unicode. I cellulari ed i computer odierni supportano Unicode, ma i primi 128 caratteri dell’enorme tabella Unicode (comprendente oggi più di 140.000 caratteri) coincidono con quelli di ASCII (che appunto ne comprendeva solo 128, ovvero 2⁷). Ad esempio, la *a* minuscola corrisponde al numero 97 in entrambe.

— ad esempio, le ricerche su Google usano algoritmi straordinariamente evoluti, che superano molti problemi di *spelling* e codifica — ma non tutte le tecnologie sono disponibili in tutti i contesti. Su un cellulare è molto più comodo digitare in ASCII che in arabo.¹⁶

L'impatto politico

Sarebbe difficile argomentare che questa 'colonizzazione tecnologica' della scrittura sia frutto di una pianificazione ostile contro le culture non-occidentali. Ma ugualmente essa ha implicazioni politiche, in quanto la scrittura, e la concezione della lingua ad essa retrostante, costituiscono aspetti centrali di molte culture, spesso sentiti come identitari. Il caso dell'arabo è più che evidente.

Poche scritture sono state oggetto e strumento di un conflitto politico così evidente come quello che si è sviluppato intorno alla scrittura araba all'inizio del Novecento nell'Impero ottomano ed ha portato all'adozione della grafia latina in Turchia nel 1928. In quegli anni lo stato turco si è costituito sul modello europeo: nazionale, laico, tanto desideroso di marcare la propria identità etnica e linguistica da 'de-arabizzare' la lingua ottomana, già mescolanza di turco, arabo e persiano, e da adottare la scrittura latina per aderire meglio, a detta dello stesso Atatürk, alle specificità fonetiche della lingua turca, in cui, peraltro, le vocali hanno un ruolo morfologico più pesante rispetto alla lingua araba.¹⁷

Di converso, i popoli arabi e maghrebini, ribelli ad un impero sempre meno ottomano e sempre più turco, hanno trovato proprio nella lingua e nella scrittura araba uno dei simboli identitari in quella lotta.¹⁸

Oggi, per quanto possa sembrare sorprendente ad un occidentale, proprio la struttura sostanzialmente sillabica della scrittura araba, con la sua distinzione di status tra consonanti, vocali lunghe e vocali brevi, ha una funzione identitaria. Essa gioca un ruolo importante nel costruire l'idea dell'unità della cultura dei popoli di lingua araba e tradizione islamica nel tempo — dal Corano ad oggi — e nello spazio — dal Marocco all'Iraq.

¹⁶Una riflessione profonda su «pratiche» scritte non occidentali e tecnologie digitali è in A. Perri, «Al di là della tecnologia, la scrittura. Il caso Unicode», art. cit, pp. 725-734. È molto interessante anche la *pars construens* del saggio alle pp. 747-748.

¹⁷J. R. Osborn, *The type of calligraphy*, op. cit, p. 5: «In 1928, Turkey adopted Latin letters for the written representation of modern Turkish. This movement of alphabetic reform was highly symbolic and fractured the connection between the textual products of Turkey and Arab tradition»; e p. 39: «The Turkish shift from Arabic to Latin letters, for example, was a symbolic statement. With the adoption of Latin-derived letters, Turkish secularism broke with Islamic tradition and visually aligned itself with Europe as a sign of modernity» (sul tema vd. anche le pp. 246-253).

¹⁸*Ibid.*, p. 255: «Contra Ottoman Turkish, celebration of the Arabic language became an anti-colonial symbol of Arabic pride and regional independence». In generale sul rapporto tra scrittura e ideologia si veda G. R. Cardona, *Antropologia della scrittura*, op. cit, par. 3.6, con due esempi relativi alla grafia araba, ma ad altre epoche storiche (vd. le pp. 125-127 sull'Uganda tra Otto e Novecento, e pp. 128-129 sulla dominazione araba della Sicilia).

Come già accennato, nella percezione di chi parla l'arabo e nella sua tradizione grammaticale, consonanti e vocali lunghe bastano a identificare per iscritto la radice della parola, mentre la parte della morfologia demandata alle altre vocali è demandata alla competenza del lettore.¹⁹

Ma c'è di più in gioco: i grafemi di consonanti e vocali lunghe di una parola — cioè la parte che viene normalmente scritta — rimangono sostanzialmente uguali nel tempo, cioè non variano significativamente dall'arabo classico del Corano al 'Modern Standard Arabic' (MSA), e nello spazio, cioè tra le aree geografiche. Al contrario l'effettiva pronuncia delle vocali brevi, normalmente tenute fuori dalla scrittura, è variabile nel tempo e nelle varianti locali della lingua.

L'arabo fuṣḥā, antico e moderno, arricchito oggi di neologismi ma conservatore rispetto alle strutture grammaticali, è l'unica lingua unificante del mondo arabo, all'interno del quale i cosiddetti vernacoli locali (tunisino, egiziano, saudita etc.) differiscono anche molto tra di loro, ad esempio nella pronuncia e nel lessico. Essendo il fuṣḥā la lingua del Corano, esso è percepito come un elemento identitario fondamentale dalle culture che si identificano con la tradizione culturale e religiosa dell'Islam, anche se di lingua diversa (pashtun, persiana etc.). E il legame con questi elementi identitari (lingua, scrittura, religione) è sempre più forte nelle società mediorientali che si percepiscono come assediata dalle pressioni globalizzanti e da condizioni socio-economiche spesso opprimenti.²⁰

Una parola in fuṣḥā è la stessa parola nel passato del Corano e nel presente, a Est come ad Ovest, cioè in tutta la storia della cultura araba (per lo più identificata con la storia dell'Islam), e in tutto lo spazio geografico della Umma (la comunità delle società di tradizione islamica), finché se ne scrivono solo le consonanti e le vocali lunghe, cioè solo all'interno del sistema grafico arabo. Scrivere la vocalizzazione, o meglio pertinentizzarla, cioè renderla significativa nel sistema grafico, significa rompere l'unità culturale (e quindi sociale, e potenzialmente politica) del mondo arabo-islamico.

Se le tecnologie digitali fossero state sviluppate originariamente in Arabia Saudita, in Israele o in India, c'è da credere che la modellizzazione della scrittura avrebbe tenuto conto anche della struttura specifica delle lingue semitiche, cioè codificando, almeno per alcuni sistemi grafici, alcuni grafemi come strutturali

¹⁹Vd. sopra il paragrafo *Le ḥarakāt arabe*.

²⁰Un esempio di questa percezione è in ROGHAYEH ROSTAMPOUR, *Footprints of the «West» in Arabic language and literature*, in Claude Alphonso Alvares e Shad Saleem Faruqi (a. c. di), *Decolonising the university: The emerging quest for non-eurocentric paradigms*, Glugor, Pulau Pinang, Penerbit Universiti Sains Malaysia, 2012: «The tendency towards vernacular and local languages in Arab countries and the explicit and implicit attempts at colonisation for the use of such local languages in academic, national, and even international conferences and meetings indicate the Western fear of Muslim unity. With the weakening of the Arabic language, Islamic-Arabic identity has also lost its essence»; la conclusione dello studioso iraniano assume toni molto forti, su cui tornerò nel mio paragrafo conclusivo: «the alienation of local languages from a familiar Qur'anic language is a plot to avoid the development of Islamic knowledge and the spread of Eastern thought and culture».

(consonanti e vocali lunghe nella scrittura araba, le consonanti nella scrittura ebraica o devānāgarī) e altri come diacritici modificatori di quelli.

Un modello di codifica di questo genere è senz'altro possibile, dal punto di vista tecnologico. Ad esempio, nella mia edizione del *De nomine* di Orso Beneventano ho usato il markup testuale TEI XML per codificare la differenza di status tra grafemi-base e diacritici abbreviativi nel sistema grafico manoscritto del latino medievale.²¹

L'attuale stadio evolutivo delle tecnologie digitali ci permette di creare modelli abbastanza complessi da rappresentare in modo non occidentalizzante oggetti culturali come le scritture 'non gutenberghiane'. Basti vedere come le tecnologie di sintesi e riconoscimento vocale hanno 'domato' in buona parte il rapporto straordinariamente complesso tra fonetica e ortografia in inglese:²² la tecnologia c'è — se c'è l'interesse strategico. La questione non è tecnologica, ma politica.

Postilla: speranze

Una postilla: in queste considerazioni finali mi sono concentrato sulla cultura araba. In essa le reazioni al profondo disagio sociale e politico prendono a volte, più spesso di quanto vorremmo, le forme del fondamentalismo religioso e della violenza. Non intendo ignorare la questione, per quanto non sia questa la sede per esaminarla a fondo. Personalmente, se da un lato condanno nettamente gli esiti fondamentalisti e quelli violenti, dall'altro penso che la difesa di un'identità e di unità culturale araba sia una reazione comprensibile di fronte al percepito sfruttamento occidentale.

Lo scenario ideale, dal mio punto di vista, sarebbe quello di un Medio Oriente socio-economicamente florido e politicamente libero, non sfruttato nello scenario della globalizzazione e non oppresso da regimi locali complici. Un Medio Oriente che conservi la propria identità culturale, inclusa la specificità della propria lingua e quindi della propria scrittura, nel tempo e nello spazio, senza bisogno di fare di quell'identità un'arma di difesa o di offesa.²³

²¹Vd. sopra il paragrafo *Per Raul Mordenti*.

²²Per un inquadramento generale dei problemi legati all'ortografia inglese, vd. G. Sampson, *Writing systems, op. cit.*, pp. 194-213 e A. Baroni, *La grafematica: teorie, problemi e applicazioni, op. cit.*, pp. 98-105. Non si dimentichi che sullo sfondo delle tecnologie che ci permettono di 'conversare' oralmente con l'assistente virtuale del nostro cellulare o dell'automobile ci sono algoritmi che trasformano le onde sonore dei nostri comandi in una sequenza di caratteri (riconoscimento vocale) e viceversa (sintesi), e queste tecnologie sono state perfezionate innanzitutto per l'inglese, nonostante esso abbia una ortografia molto 'opaca'.

²³Il tema dell'intersezione tra cultura e politica nei rapporti tra "Oriente" e "Occidente", tra "Nord" e "Sud" globali, coincidente in sostanza con l'area nota come "studi postcoloniali", è naturalmente troppo ampio per affrontarlo qui. Rimanderò solo ad un libro che si può considerare l'apripista di questo ambito di studi, EDWARD WADIE SAID, *Orientalism*, London, Routledge & Kegan Paul, 1978.