Note di lessicografia

di ANNARITA PUGLIELLI (Roma)

La redazione di un dizionario può apparire un'operazione di routine, addirittura resa banale oggi dal fatto che ci si può servire di supporti informatici che consentono sia di maneggiare una quantità di dati una volta impensabili, sia di fare tutta una serie di operazioni – quali controlli incrociati, ecc. – che sarebbe impossibile fare manualmente. Questo da una parte consente l'uso di *corpora* immensi[1] in cui sono registrati, e quindi resi disponibili, usi provenienti da diversi contesti (lingua scritta e lingua parlata, registri e stili diversi, aree e campi diversificati ecc.) e dall'altra, in qualche modo, stimola la produttività, sia per quanto riguarda i dizionari monolingue che quelli bilingue. Si assiste così, anche in contesti come quello italiano in cui la tradizione lessicografica risale al XVI secolo ed ha sicuramente dato numerosi e ampi risultati, alla produzione continua di nuovi dizionari.

Ma in che senso questi dizionari possono dirsi nuovi? Sicuramente la novità riguarda la selezione dei lemmi inclusi, con la registrazione di parole che entrano nell'uso e si impongono diventando così "di moda", ma hanno qualcosa di nuovo, di diverso anche nella loro organizzazione strutturale e/o concettuale?

Se ripercorriamo in modo assai rapido la storia della nostra lessicografia ci si rende conto che si giunge a quello che ormai tendiamo a considerare il formato standard delle voci lessicali attraverso approssimazioni successive. Così, ad esempio, in uno dei primi vocabolari alfabetici italiani dal secolo XVI, quello del Luna (1536), ciò che colpisce un lettore di oggi è la asistematicità. Non coerenti sono le entrate lessicali per i verbi che vengono citati sia al passato che al pre-

[1] Si veda ad esempio il *Collins Cobuild English Language Dictionary* per la messa a punto del quale è stato utilizzato la Birmingham Collection of English Texts, un *corpus* costituito da molti milioni di parole provenienti da libri, periodici, quotidiani, libelli, saggi, conversazioni e trasmissioni radio e televisive.

sente e spesso, ma non sempre, alla prima persona; la classificazione grammaticale delle voci appare raramente e non in modo sistematico. Nella descrizione vengono forniti essenzialmente esempi d'uso tratti dai classici (Dante, Petrarca e Boccaccio) e solo a volte vengono inseriti a titolo esplicativo dei sinonimi introdotti da cioè. Si tratta dei primi tentativi di redazione di vocabolari monolingue – dato che fino ad allora si erano prodotti dizionari bilingue o plurilingue che ovviamente pongono problemi diversi – e quindi l'elaborazione del sistema era proprio nella fase iniziale.

Ben altro livello di elaborazione si ritrova invece nel *Vocabolario degli Accademici della Crusca* (iniziato nel 1591 e pubblicato nel 1612) che ha rappresentato per diversi secoli il lessico modello. Nella prefazione alla V edizione stampata a Venezia nel 1741 vengono elencati così i punti fondamentali dell'organizzazione del *Vocabolario*:

I) scelta delle voci (per cui non vengono esplicitati i criteri);

II) definizione di ciascuna voce, che deve essere il più breve possibile. Si dice esplicitamente che a volte invece della definizione viene data «una voce della stesso significato comunalmente detta "sinonimo"»;

III) la definizione o spiegazione della voce è seguita dall'equivalente latino e greco;

IV) ciascuna voce è accompagnata da «una copiosa ricchezza di esempi»;

V) vengono anche forniti significati «men propi» oppure, come contesti, proverbi particolari;

VI) non sono state introdotte le etimologie;

VII) sono forniti tutti gli indici necessari degli autori, delle abbreviazioni, ecc.;

VIII) particolare cura è rivolta agli aspetti ortografici.

Questa dunque l'impostazione esplicitata; possiamo aggiungere che la forma di citazione dei verbi adottata è l'infinito, e che la categoria grammaticale è presente solo in alcune voci (preposizioni, avverbi e comparativi). Inoltre sono anche presenti alcune informazioni morfologiche di carattere derivazionale (per es. indicazione di prefissi: *abbisognoso da bisognare* ecc.).

La descrizione fornita, anche se coglie solo gli aspetti fondamentali, è sufficiente a mostrare come il *Vocabolario degli Accademici della Crusca* rappresenti una prima e senz'altro avanzata formalizzazione di quello che con qualche variante e completamento diventerà poi il formato standard di un vocabolario.

Gli approfondimenti riguarderanno fondamentalmente tre aspetti: *a*) l'introduzione sistematica della categoria grammaticale in ogni voce del dizionario, *b*) l'introduzione per alcune voci di una classificazione semantica, *c*) l'accentuazione – a partire dal XIX secolo – della regi-

strazione documentaria dell'uso della lingua nella sua totalità piuttosto che del buon uso tratto da fonti classiche.

Così, attraverso la pratica di secoli, si è giunti a una formulazione delle voci lessicali che si può considerare ormai nella nostra tradizione occidentale standardizzata e che possiamo schematizzare così: lemma, categoria grammaticale, informazione morfologica, classificazione semantica (cioè specificazione del campo concettuale di appartenenza del lemma o del particolare significato), definizione di ciascun significato del lemma, sinonimi, esempi di contesti d'uso tratti da fonti diversificate.

Naturalmente a questo nucleo di informazioni che può essere a diversi livelli di approfondimento, se ne aggiunge altra che è spesso, ma non sempre, presente: la trascrizione fonetica, le varianti (nella lingua standard e/o regionalismi ecc.) i contrari. Nel *Collins Cobuild* (1987) è stata anche introdotta una colonna in più in cui viene data informazione grammaticale o sintattica e informazione semantica[2].

Nella nostra tradizione lessicografica la compilazione di un dizionario può dunque apparire un'operazione scontata e che necessita di decisioni solo sulla quantità di informazione da includere tra le varie possibili e ormai standard, mentre la compilazione di un dizionario diventa immediatamente meno ovvia quando ci si trova a scrivere di fatto il primo dizionario sistematico di una lingua. In questo caso ci si incontra, o meglio scontra, non solo con una serie di problemi specifici determinati dalle caratteristiche della lingua oggetto della descrizione, ma anche con problemi di ordine generale e teorico che riguardano proprio i contenuti e l'organizzazione delle entrate lessicali. In altri termini alcuni aspetti che sembravano ormai definitivi vengono rimessi in discussione proprio dalla necessità di far fronte ad esigenze in qualche modo «nuove». L'impatto di natura contenutistica e metodologica creato dal nuovo contesto poi, quasi per induzione, porta a riinterrogarsi almeno su alcuni di quegli aspetti che appaiono ormai come consolidati e stabili per le lingue occidentali di lunga tradizione lessicografica.

Sarebbe ben al di là di queste brevi note un'analisi ragionevol-

[2] L'informazione semantica non costituisce di per sé una novità ma ai sinonimi sono stati aggiunti i contrari e gli iperonimi. L'informazione grammaticale non si limita a fornire la categoria grammaticale, ma fornisce anche le sottocategorizzazioni all'interno di ciascuna categoria e il contesto grammaticale in cui si può trovare. Così ad esempio, gli aggettivi sono sottoclassificati in aggettivo classificativo, aggettivo che segue N, aggettivo-colore, aggettivo qualitativo, e inoltre aggettivo attributivo (per aggettivi che precedono sempre il nome) e aggettivo predicativo (per aggettivi che di solito occorrono solo dopo un verbo come *be*, *seem*, *become*: es. *asleep* in *the baby is asleep* ma non *the asleep baby*.

mente esauriente delle tematiche e problematiche che si sente di dover rimettere in discussione e che vanno da temi di fondo quali il rapporto tra teoria linguistica (grammaticale semantica ecc.) e prassi lessicografica, ad elementi di dattaglio come ad esempio secondo quali criteri ordinare i diversi significati all'interno di una voce lessicale, o quali e quante varianti includere (problema di particolare difficoltà ad esempio per lingue che solo di recente hanno adottato un sistema di trascrizione, trasformandosi da lingue orali in lingue scritte).

Nell'impossibilità dunque di trattare esaurientemente tutti i temi pertinenti, intendiamo in questa sede limitarci a qualche riflessione su due degli aspetti che nella costruzione dell'entrata lessicale sembrano più che consolidati: l'attribuzione della categoria grammaticale e l'esplicitazione dell'area semantico-concettuale di appartenenza di una determinata voce.

Per quanto riguarda le categorie grammaticali, l'esperienza di lavoro con una lingua come il somalo[3] ci ha fatto riconfrontare con antichi dilemmi come quali e quante categorie grammaticali utilizzare e quindi in base a quali tipi di criteri esse vengano determinate. Saranno da utilizzare criteri morfologici o sintattico-semantici? E poi, una volta presa una decisione su base teorico-descrittiva come dovrà rispecchiarsi questa nel dizionario? Gli aggettivi, che nella identificazione delle parti del discorso sono stati introdotti solo in un momento successivo essendo la prima distinzione fatta quella tra nomi/sostantivi e verbi, sono uno dei punti più discussi e controversi nella identificazione delle parti del discorso in somalo e dell'attribuzuione ad esse delle parole della lingua[4].

Infatti mentre Reinisch (1903) introduce nella sua descrizione la categoria degli aggettivi – con sottoclassificazioni di tipo morfologico in aggettivi primitivi ed aggettivi derivati – applicando in modo piuttosto automatico alla descrizione del somalo categorie necessarie per la descrizione di lingue occidentali, Bell (1953) si è reso conto che la definizione di una classe di aggettivi e l'attribuzione ad essa delle parole pertinenti era un problema estremamente complesso, e, pur mantenendo questa categoria considerò aggettivi solo quelle parole da cui si possono formare verbi attributivi. Solo gli approfonditi studi dei paradigmi verbali condotti successivamente da Andrzejewski (1956, 1969) hanno mostrato che quelli che Bell considerava aggettivi altro non erano che forma del presente di un particolare paradigma verbale, il

[3] Il somalo è una lingua del gruppo cuscitico, della famiglia afroasiatica (cfr. Abdalla O. Mansur 1989). Per dati sulla ricerca svolta su questa lingua soprattutto a livello lessicografico si veda Puglielli (1994).

[4] Per una dettagliata descrizione di questo aspetto si veda Banti (1994).

cosiddetto 'paradigma ristretto' che compare in somalo ogni volta che il soggetto della frase è focalizzato, cioè marcato come informazione nuova dalla particella *baa* (Puglielli 1981: 13).

Questo e altri argomenti di tipo sintattico come la necessità di ricondurre strutture nome + modificatore a intere frasi per arrivare a formulare generalizzazioni possibili, nonché la struttura della frase relativa, hanno portato a considerare le parole che nella prima descrizione erano state considerate aggettivi, una classe di verbi attribuiti alla quarta coniugazione, ovvero, come li ha denominati Andrzejewski (1969) «verbi ibridi». Da un punto di vista semantico i verbi della quarta coniugazione sono stativi, cioè hanno le stesse caratteristiche degli aggettivi nelle nostre lingue[5].

Che riflesso ha avuto tutto ciò nella compilazione del *Dizionario Somalo Italiano* prima e del *Grande Dizionario Somalo Monolingue*[6] poi? Se per omologazione alle nostre lingue e sulla base di caratteristiche semantiche, poteva apparire plausibile considerare parole come *xun* "brutto", *dheer* "piccolo", *ladan* "ricco", *weyn* "grande" come aggettivi, sulla base del sistema complessivo della lingua e di considerazioni di carattere prevalentemente morfologico-sintattico tale scelta appariva non sifficientemente giustificata. Da qui la decisione di introdurre quattro coniugazioni di verbi e accorpare in essa anche quelli più tradizionalmente considerati aggettivi.

Casi come questi o simili portano a rimettere in discussione dunque non solo la determinazione di quali siano le categorie grammaticali necessarie per la descrizione di una determinata lingua, ma anche su quali criteri ci si dabba basare per arrivare alla determinazione di tali categorie e all'attribuzione delle parole a ciascuna di esse.

Questo tipo di problemi così evidenti quando si ha a che fare con lingue "diverse", potrebbe apparire superato nella nostra tradizione lessicografica, ed invece non è ancora del tutto risolto. Per un esempio rapido si pensi al trattamento nei dizionari italiani di parole come *letto*, *litigato*, *contestato*. Si tratta del participio passato di alcuni verbi che, contrariamente alle altre forme del verbo, viene citato come una voce separata rispetto a quella verbale dato che può occorrere oltre che nei tempi composti insieme all'ausiliare, anche con funzione aggettivale, cioè come modificatore di un nome. Queste parole dunque vengono

[5] La questione degli aggettivi in somalo è ancora tema di dibattito: cfr. Saeed (1988), Ajello e Puglielli (1988), Banti (1988), Banti (1994).

[6] Il materiale del *Dizionario Somalo Monolingue*, la cui redazione è stata portata avanti in Somalia dal 1986 al 1990, era fortunatamente stato portato in copia in Italia prima dell'inizio della guerra civile, ed è attualmente in fase di revisione a Roma presso il Dipartimento di Linguistica della Terza Università degli Studi.

categorizzate in due modi diversi, come verbi e come aggettivi[7]. Va però notato che la doppia funzione e quindi la possibilità di categorizzare una parola in due modi diversi non è sempre considerata elemento sufficiente per arrivare ad avere entrate lessicali separate (ad esempio non è classificato in modo diverso un aggettivo attributivo e un aggetivo predicativo). Nel caso del participio passato la forma che è morfologicamente diversa dalla forma di citazione del verbo può essere motivo sufficiente per motivare questa scelta, tuttavia una volta fatta, tale scelta dovrebbe essere applicata coerentemente in tutti i casi. Questo non è sempre vero; per esempio nel Dizionario Zingarelli ci sono le seguenti voci: *letto*, *contestato*, *deformato*, ma mancano voci come *arroccato*, *disgregato*, *defraudato*.

Sempre riflettendo sul problema dell'attribuzione di lemmi alle categorie grammaticali, attualmente nella nostra tradizione lessicografica la classificazione in nome, verbo ecc. appare irrinunciabile, ma se in una lingua si trova un numero assai numeroso di lemmi che hanno esattamente la stessa base e possono essere usati sia come nomi che come verbi, che possono essere cioè inseriti sia in un contesto sintattico di tipo nominale (per esempio possono essere preceduti dall'articolo) e assumere quindi marche morfologiche di tipo nominale come plurali o casi, ovvero essere inseriti come predicato nella frase, assumendo la flessione temporale e modale del verbo, non sarebbe più che motivata una scelta che riflettesse questa caratteristica presentando una sola entrata lessicale con la forma radicale o di base della parola evidenziando così questa regolarità?[8]

Un altro aspetto cui vogliamo accennare, e che appare ancora non risolto anche nella nostra tradizione lessicografica, è quello della specificazione del "campo semantico" o area concettuale di appartenenza delle parole. Consultando i vari dizionari anche più recenti che potremmo chiamare 'generali' – cioè non dedicati a un particolare settore disciplinare, come ad esempio medicina, biologia, chimica ecc. – si nota che è ancora nella maggior parte dei casi presente per alcuni lemmi la specificazione dell'appartenenza ad un particolare settore come ad esempio architettura, medicina, chimica, geografia ecc. Non sono però esplicitati né i criteri in base ai quali tale specificazione viene inserita – dal momento che non è inserita in tutti i lemmi classificabili come appartenenti a una determinata area – né tanto meno la

[7] Nell'entrata lessicale dei dizionari l'ordine di presentazione seguito è participio passato, aggettivo. Tale ordine pur rispondendo a criteri di praticità per il rinvio immediato al verbo, non risponde a criteri di coerenza del sistema che vorrebbe la categoria aggettivo in prima posizione dato che è quello l'uso che determina una voce separata.

[8] Questo modo di procedere è peraltro usato nella lessicografia tradizionale di lingue come l'arabo e non costituisce quindi una novità; si tratterebbe di estenderne l'uso anche ad altre lingue di altre famiglie.

griglia di riferimento, il sistema usato per tale classificazione. Così, accanto a campi di una certa generalità come ad esempio medicina, se ne trovano altri di livello più specifico come ad esempio chirurgia; si ritiene necessario specificare quindi un particolare settore della medicina ma non altri quali potrebbero essere cardiologia, ortopedia ecc., né è mai esplicitato in base a quali criteri queste scelte vengono operate. L'impressione che se ne ricava è che per motivi di economia generale del dizionario queste classificazioni vengono di solito inserite in voci lessicali in cui il lemma in questione è utilizzato con significati diversi in aree semantiche diverse; ne è un esempio la parole italiana *timpano* che è utilizzato come termine anatomico, musicale e anche in architettura. Tuttavia anche questa impressione non trova conferma, infatti nell'entrata lessicale della parola *vascolare* la specifica "anatomia" è presente pur essendo per quel lemma registrato solo quell'uso e l'interpretazione ad esso legata[9].

D'altro canto la classificazione più che da una riflessione di tipo teorico sull'identificazione di tassonomie sistematiche delle aree semantiche da utilizzare, sembra essere determinata su basi puramente empiriche e intuitive in rapporto a un'idea latente di «conoscenze del parlante di media cultura». Una informazione dunque che appare necessaria dal confronto con i dati, ma a cui non si fornisce una risposta adeguata a livello teorico.

Questo aspetto appare ovviamente ancora più complesso quando il dizionario deve fornire dati su una lingua per diversi aspetti lontana da quelle occidentali su cui la nostra tradizione lessicografica si fonda. In questi casi dunque volendo introdurre classificazioni semantiche, su quali criteri ci si dovrà basare? Sarà opportuno cercare di identificare solo aree semantiche "generali" e quindi applicabili trasversalmente a tutte le lingue, oppure identificarne anche altre più interne all'ambito culturale della lingua descritta? In altri termini, le acquisizioni di ambiti di studio di recente sviluppo quali ad esempio l'etnoscienza, quanto potranno o dovranno essere tenute in conto nell'elaborazione di dizionari di lingue cosiddette "esotiche"?

I temi che abbiamo toccato più che trattato, come altri che per mancanza di spazio non abbiamo neanche menzionato, hanno bisogno, a nostra avviso, ancora di approfondimento e riflessione; solo così si potrà forse arrivare alla compilazione di dizionari "nuovi".

[9] Il vocabolario cui ci riferiamo è quello di N. Zingarelli (1994).

AA.VV. (1987), *Collins Cobuilt English Language Dictionary*, London-Glasgow.
Abdalla O. Mansur (1989), *Le lingue cuscitiche e il somalo, Studi Somali 8*, (edizione italiana a cura di C. Serra Borneto), Roma.
Agostini, F., Puglielli, A., Ciise Mohamed Sijaad, (1985), *Dizionario Somalo-Italiano*, Roma.
Ajello, R., Puglielli, A., (1988), *More on "hybrid verbs" and other categories in Somali*, in Bechhaus-Gerst, M., Serzisko, F. (eds) *Cushitic-Omotic*, Amburgo, pp. 535-551.
Andrzejewski, B.W., (1956), *Accentual patterns in verbal forms in the Isaaq dialect of Somali*, in «Bullettin of the School of Oriental and African Studies», 18, pp. 103-129.
–, (1969), *Some observations on hybrid verbs in Somali*, in «African Language Studies», 9, pp. 1-51.
Banti, G. (1988), *Adejctives in East Cuschitic*, in Bechhaus-Gerst, M., e Serzisko, F. (eds), *Cushitic-Omotic*, Amburgo, pp. 203-259.
–, (1994), *Parti del discorso controverse*, in *Lingua, pensiero scientifico e interculturalità*, Atti dei Convegni dei Lincei 107, Roma, pp. 51-82.
Bell, C.R.V. (1953), *The Somali Language*, Londra.
Luna, F. (1536), *Vocabolario di cinquemila vocaboli toschi*, Napoli.
Migliorini, B. (1946), *Che cos'è un vocabolario?*, Roma.
Pitteri, F. (ed.) (1741), *Vocabolario degli accademici della Crusca*, V Impressione, Venezia.
Puglielli, A. (1981a), *Frase dichiarativa semplice*, in Puglielli, A. (ed.), *Sintassi della lingua somala*, Roma.
– (1981b), *Sintassi della lingua somala*, Studi Somali 2, Roma.
– (1994), *Il progetto sulla lingua somala: obiettivi e metodi*, in *Lingua, pensiero scientifico e interculturalità*, Atti dei Convegni dei Lincei 107, Roma, pp. 31-44.
Reinisch, L. (1903), *Die Somali Sprache*, III (Grammatik), Vienna.
Saeed, J. (1988), *An Argument for the category Adjective in Somali*, in Bechhaus-Gerst, M., Serzisko, F. (eds), *Cushitic-Omotic*, Amburgo.
Zingarelli, N. (1994), *Vocabolario della lingua italiana*, XII edizione a cura di M. Dogliotti, L. Rosiello, Bologna.