

4. PROCEDURA ELETTRONICA DI SPOGLIO.

Le elaborazioni per la produzione della concordanza dei grammatici latini sono state eseguite applicando un sottoinsieme del sistema per lo spoglio dei testi in linguaggio naturale attualmente disponibili presso il Laboratorio di Linguistica computazionale del CNR.

Questo sistema (1) può essere considerato come un generatore di procedure generalizzate (general purpose) nel settore del text-processing. La sua generalità deriva essenzialmente da due caratteristiche:

a) esso consiste in una serie di programmi parametrizzati, i quali, applicati secondo diverse combinazioni, possono produrre l'intera gamma dei risultati normalmente richiesti come 'output' di uno spoglio: concordanze di vario tipo (integrali, selettive, contrastive, sintattiche, tematiche (2), ecc.); indici diversi (index verborum, laterculum, rationarium, rimari, incipitari, indici inversi, indici di frequenza, ecc.); schede contestualizzate, ecc.;

b) esso comprende uno standard per la rappresentazione di testi di epoche, lingue, generi diversi.

I programmi sono parametrizzati nel senso che le specifiche esigenze relative ai singoli casi di impiego vengono soddisfatte mediante l'impiego di parametri che li adattano a tali esigenze.

Non è questo il luogo per illustrare in dettaglio la flessibilità dei diversi programmi. Ci limiteremo a un esempio, quello della contestualizzazione, per dare un'idea della varietà delle scelte operative che l'utente può compiere, sia per quanto riguarda la organizzazione e la forma dei risultati, sia per quanto riguarda le regole algoritmo-linguistiche applicate per produrli.

(1) Il sistema in questione è stato messo a punto negli anni 1964 - 65 da A. Zampolli per gli spogli dell'Index Thomisticus e dell'Opera del Vocabolario dell'Accademia della Crusca. Successivamente, esso è stato implementato sui calcolatori del CNUCE dagli analisti-programmatori della Divisione Linguistica del CNUCE, recentemente trasformata in Laboratorio di Linguistica Computazionale del Comitato 08 del CNR. Il sistema è stato arricchito progressivamente sulla base delle esperienze condotte in collaborazione con i numerosi Istituti italiani e stranieri che lo hanno via via utilizzato. Le ricerche attualmente in corso hanno lo scopo di automatizzare le operazioni di analisi linguistica più frequentemente eseguita nelle procedure di spoglio (in particolare la lemmatizzazione), di rendere il sistema completamente interattivo, e di conferire agli archivi lessicali le caratteristiche e le funzioni previste dalle più recenti metodologie dei «data base».

(2) Per una descrizione di questi diversi tipi di concordanze si veda A. Zampolli, *La section linguistique du CNUCE*, in *Linguistica matematica e calcolatori*, a cura di A. Zampolli, Firenze, 1973, pp. 133 - 199.

L'utente che vuole produrre una concordanza può:

1) specificare quali unità linguistiche porre in esponente (lemmi; lemmi e forme; forme lessicali; forme grafiche; sintagmi; strutture sintattiche; codici semantici o tematici; ecc.);

2) specificare gli elementi che costituiscono l'esponente (per es. nel caso del lemma parola-lemma, codici morfosintattici, codici di omografia, eventuali commenti e rinvii) e le modalità di formulazione (naturalmente entro un certo limite di caratteri);

3) scegliere tra spoglio integrale e selettivo: portare in esponente tutte le unità del livello prescelto, o solo alcune (per es. vengono spesso omesse le parole vuote o grammaticali); per ciascun esponente, riportare i contesti di alcune occorrenze opportunamente prescelte, tralasciando le altre o riportandone solo i riferimenti;

4) decidere l'ordinamento dei diversi elementi della concordanza, sia per quanto concerne la sequenza degli esponenti (per es., se disporre i lemmi e le forme sotto i relativi lemmi in una unica serie alfabetica, o divisi per categoria grammaticale, o per lingua, ecc.), sia per quanto riguarda l'ordine dei contesti sotto i relativi esponenti (in ordine di testo, in ordine cronologico, ordinati secondo le parole che seguono e/o precedono nel contesto, ecc.);

5) scegliere le regole per il 'taglio' dei contesti, in un insieme di regole disponibili, che abbiamo messo a punto sulla base di un inventario dei criteri di contestualizzazione più frequentemente adoperati. Questi criteri possono essere raggruppati nei tipi seguenti:

a) Nei sistemi di kwic-index, messi a punto soprattutto per applicazioni documentarie, tutte le parole dei titoli che compongono la lista bibliografica da elaborare, o, più spesso, le sole parole 'lessicali' che vi compaiono, vengono elencate in ordine alfabetico, e ricevono come contesti i titoli, o parti di titolo, nei quali occorrono. I programmi di kwic-index oggi più diffusi non sono adeguati per compilare concordanze a scopo lessicografico, per diversi motivi.

b) La parola esponente è sempre al centro del suo contesto, ossia è sempre preceduta e seguita da un egual numero di battute o di parole. Il programma è di semplice stesura e la composizione dei contesti velocissima; spesso alla scelta di questo metodo si accompagna la decisione di non lemmatizzare, nell'evidente proposito di sfruttare al massimo la velocità della macchina e di eliminare ogni intervento umano, ed i contesti vengono ordinati, sotto le rispettive forme, secondo l'ordine alfabetico delle parole che seguono (e/o precedono) la parola esponente nel suo contesto.

c) Il contesto è costituito da un'intera unità di riferimento: il verso, il paragrafo, il versetto, il comma, ecc. Ovviamente il contesto è tanto più 'significativo' quanto più l'unità di riferimento è correlata a una unità di natura linguistica; a tale correlazione è inversamente proporzionale il rischio che il contesto sia inadeguato, soprattutto se la parola esponente occorre ai limiti del contesto stesso.

d) I limiti di contesto sono segnati in fase di 'preedizione': il testo viene suddiviso in 'pericopi' per mezzo di contrassegni che vengono perforati e conservati nelle successive elaborazioni: ciascuna pericope funge da contesto per tutte le parole che la compongono.

e) Il contesto è costituito sempre e solo da tutte le parole comprese tra due segni di punteggiatura. I tipi c, d, e hanno in comune una caratteristica ben precisa: il testo è segmentato in 'sintagmi' successivi e tutte le parole di un sintagma hanno l'intero sintagma per contesto, cioè hanno il medesimo contesto.

f) Il contesto è scelto in base alla natura della parola: per le parole grammaticali è spesso un 'trinomio' del quale la parola grammaticale è al centro, per le preposizioni sono prese per lo più le due parole successive, ecc. Il presupposto è, evidentemente, che le parole di cui si deve costruire il contesto siano già, in qualche modo, classificate; quindi questo metodo è usato spesso per concordanze di lemmi o comunque nella parte conclusiva dello spoglio (e non come strumento di lavoro, per es., nella fase di lemmatizzazione). Sono interessanti a questo proposito le possibilità di automatizzare almeno in parte l'analisi sintattica, scegliendo,

per ogni parola, quella parte terminale della struttura che si giudica interessante come contesto per la categoria grammaticale a cui la parola appartiene.

g) Il calcolatore fornisce, con un metodo qualsiasi (per lo più di tipo *d*), un primo contesto spesso sovrabbondante, che viene poi ridotto alle dimensioni richieste per mezzo di schede con cui si comunicano al calcolatore le parole che lo studioso, dopo un accurato esame, ha deciso di eliminare per snellire il contesto.

h) Il contesto è regolato tenendo conto di determinati segni quali l'interpunzione, il cambio di riferimento, ecc. (3). Come nel tipo *b*, il contesto viene costruito per ogni parola, cosicché varia da una parola a quella successiva, ma, a differenza del tipo *b* e a somiglianza invece dei tipi *c*, *d*, *e*, è regolato sulla presenza di elementi ben definiti, cosicché la parola può trovarsi collocata diversamente nel contesto: verso l'inizio, verso il centro, verso la fine, a seconda dei casi.

L'algoritmo del nostro programma è potenzialmente in grado di generare contesti secondo tutti i tipi, anche se non è stato sufficientemente sperimentato il tipo *f*, dal momento che usiamo le concordanze in fase di lemmatizzazione prima di qualsiasi analisi. Per i grammatici latini è stato scelto un contesto di tipo *b* (4).

È possibile anche specificare quali elementi del testo vadano 'contestualizzati' e quali no, ed elencare e classificare gli elementi che hanno la funzione di 'limiti' di contesto.

Nel nostro programma gli elementi del testo devono inoltre essere classificati in:

a) elementi che devono avere un proprio contesto e devono essere presenti nei contesti degli altri (per es. le parole 'lessicali');

b) elementi che devono entrare a far parte dei contesti degli altri, ma non ricevono contesto proprio (per es. i segni di punteggiatura);

c) elementi che non devono né entrare a far parte dei contesti né ricevere contesto proprio; di solito si tratta di 'codici' introdotti nel testo per compiere alcune operazioni del programma (per es. i segni di divisione in pagine e in righe);

d) elementi che devono ricevere contesto proprio, ma non entrare nei contesti altrui (per es. le varianti, qualora se ne voglia tener conto, nelle elaborazioni di un testo fornito d'apparato critico).

La possibilità di applicare il nostro sistema di text-processing a lingue, epoche e generi letterari diversi presuppone uno standard per rappresentare in 'machine

(3) L'utente del nostro programma che opta per questo tipo di contestualizzazione ha la possibilità di specificare gli elementi che debbono fungere da 'delimitatori' di contesto. Questi limiti possono essere distribuiti in classi ordinate gerarchicamente, fino a un massimo di 9. Gli elementi della classe 9 sono 'limiti invalicabili' di contesto: cioè, se nel costruire la parte destra di un contesto si incontra un limite di classe 9, il contesto non procederà più in alcun modo verso destra, qualunque sia lo spazio a disposizione. Si definiscono spesso limiti di classe 9, per esempio, i contrassegni di un cambiamento di capitolo. Gli elementi delle altre classi, dalla 1 alla 8, sono 'limiti valicabili' di contesto. Se nel costruire una sola parte di un contesto si incontra uno di questi limiti, il contesto potrà scavalcarlo e riprocedere verso destra solo dopo aver trovato un limite di classe uguale o maggiore nel procedere verso la direzione opposta. Di solito, i vari segni di punteggiatura vengono attribuiti a classi diverse, secondo la loro forza: per esempio, il punto fermo, l'interrogativo e l'esclamativo alla classe 3; le parentesi, le virgolette, i due punti, il punto e virgola alla classe 2; la virgola e la lineetta di inciso alla classe 1, ecc.

(4) I motivi di questa scelta sono esposti nella nota di N. Marinone.

readable form' la straordinaria varietà di 'grafemi' che possono apparire nei testi in linguaggio naturale. (Con il termine grafemi indichiamo qui non solo i caratteri alfanumerici stampati nel testo, ma anche informazioni fornite dal testo in altra forma — spaziature, suddivisioni, ecc. — o inserite nel testo in fase di preedizione: per es., contrassegni di parti non autentiche, di citazioni di altri autori, ecc.).

In effetti, abbiamo messo a punto uno standard costruito sulla base di un inventario che tiene conto delle esperienze di spoglio nostre e dei principali centri di linguistica computazionale europei e nordamericani. Esso può essere presentato come una tabella che specifica, per ogni grafema (o, più esattamente, per ciascuno dei suoi significati) il codice che lo rappresenta nei diversi supporti previsti dal sistema (schede, nastri, diverse tecnologie di stampa, ecc.), e le sue funzioni nelle varie fasi dell'elaborazione.

Questo standard viene utilizzato praticamente da tutti i progetti di text-processing in corso nel nostro paese nel settore delle ricerche linguistiche e filologiche. Questo fatto garantisce la reciproca scambiabilità dei testi registrati in 'machine readable form' da ricercatori o istituti diversi. Per valutare appieno il significato di questa situazione, che ora anche altri paesi stanno cercando di realizzare, si deve tenere presente che, sul piano finanziario e organizzativo, la registrazione di un testo è la fase più lunga e più costosa di uno spoglio. Pertanto è opportuno che ogni testo registrato possa essere utilizzato da tutta la comunità degli studiosi. Sul piano scientifico nessuno dubita che alcune importanti dimensioni, caratterizzanti le ricerche linguistiche e filologiche, si fondano sulla raccolta dei dati, dunque sulla osservazione, e richiedono una organizzazione speciale del lavoro e in particolare la ricerca di gruppo su corpora di grandi dimensioni. La adozione di uno standard di registrazione permette che i testi registrati in luoghi e per scopi diversi confluiscono a costituire corpora omogenei sempre più rappresentativi e controllabili articolati e strutturati lungo le diverse dimensioni (storiche, di genere letterario, sociolinguistiche, ecc.) di una lingua. La presenza di questi corpora facilita e rende più economica la raccolta dei dati, e consente anche di verificare su una base adeguata modelli e ipotesi teoriche di varia natura (5).

(5) Si pensi, per esempio, allo stato attuale della statistica linguistica. I modelli e le cosiddette 'leggi' proposte sulla base degli spogli disponibili negli anni '50 sono stati largamente falsificati dai dati quantitativi forniti dagli spogli elettronici negli ultimi anni, e le ricerche in corso mirano alla formulazione di nuovi modelli capaci di rendere conto dei dati forniti dai corpora di sempre maggiori dimensioni e sempre più articolati e stratificati che vengono via via accumulandosi in 'machine readable form' (cf. A. Zampolli, *L'elaborazione elettronica dei dati linguistici: Stato delle ricerche e prospettive*, in *Colloquio sul tema: Le tecniche di classificazione e loro applicazione linguistica*, Accademia Nazionale dei Lincei, Roma 1975, pp. 23 - 107, e in particolare le pagine 58 e segg.).

In effetti, l'adozione del nostro standard ha portato alla creazione di una 'biblioteca' che è tra le più estese oggi esistenti. Essa costituisce il primo passo per la costruzione di una vera e propria banca di dati linguistici. Comprende oltre 5000 testi, registrati con uniformità di criteri tecnici e scientifici, ed elaborabili perciò con gli stessi programmi fondamentali.

Ogniqualevolta un nuovo programma viene scritto per una nuova elaborazione richiesta da un progetto specifico, esso diviene immediatamente applicabile a tutti i 'testi' registrati per altri progetti. Inoltre l'adozione di uno schema unico di registrazione evita al ricercatore che si accinge allo spoglio di un testo il lungo e difficile lavoro di stabilire le norme di perforazione, e assicura nel contempo che vengano registrate tutte le informazioni presenti nel testo a livello grafemico, e quindi ne garantisce la utilizzazione anche per ricerche successive a quella che interessa il ricercatore in questione.

Si potrebbe dire che lo standard funziona come un 'questionario' che chiede al ricercatore se un certo fenomeno, teoricamente possibile, è presente o no nel testo, e che quindi lo guida passo passo all'analisi delle informazioni da rappresentare in input e alla definizione delle loro funzioni nelle elaborazioni successive.

La procedura che stiamo seguendo per produrre gli indici e le concordanze dei grammatici latini è costituita, come si è detto, dalla concatenazione di un sottoinsieme di componenti operativi, opportunamente prescelti, del nostro sistema di text-processing.

Farà però eccezione il programma per la normalizzazione semiautomatica delle forme-esponenti delle concordanze (cfr. op. 17 qui sotto), che è stato studiato per le esigenze dei grammatici latini.

Il diagramma operativo che si riporta, suddiviso per comodità di consultazione in sette diverse tavole, rappresenta il coordinamento logico e cronologico delle operazioni eseguite. Vi sono usati i seguenti

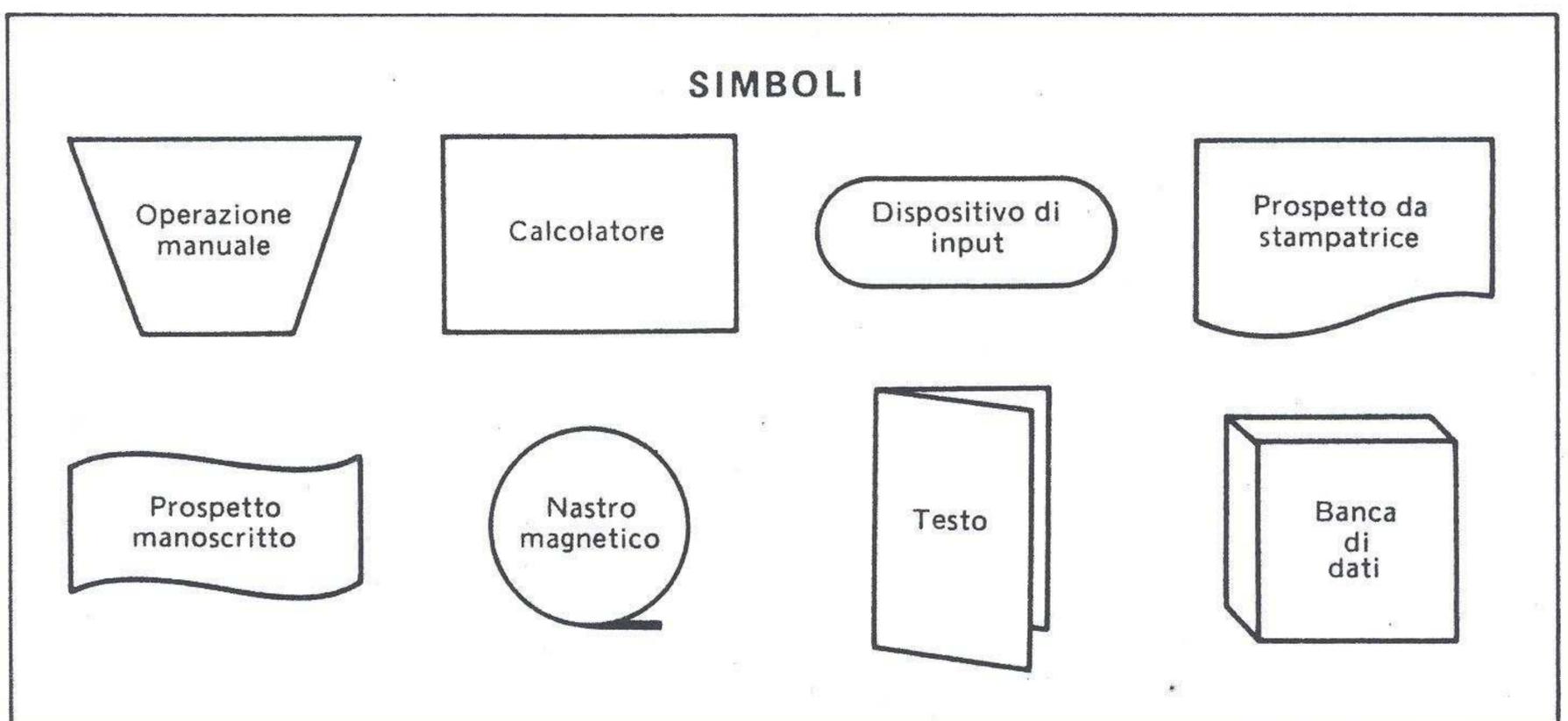
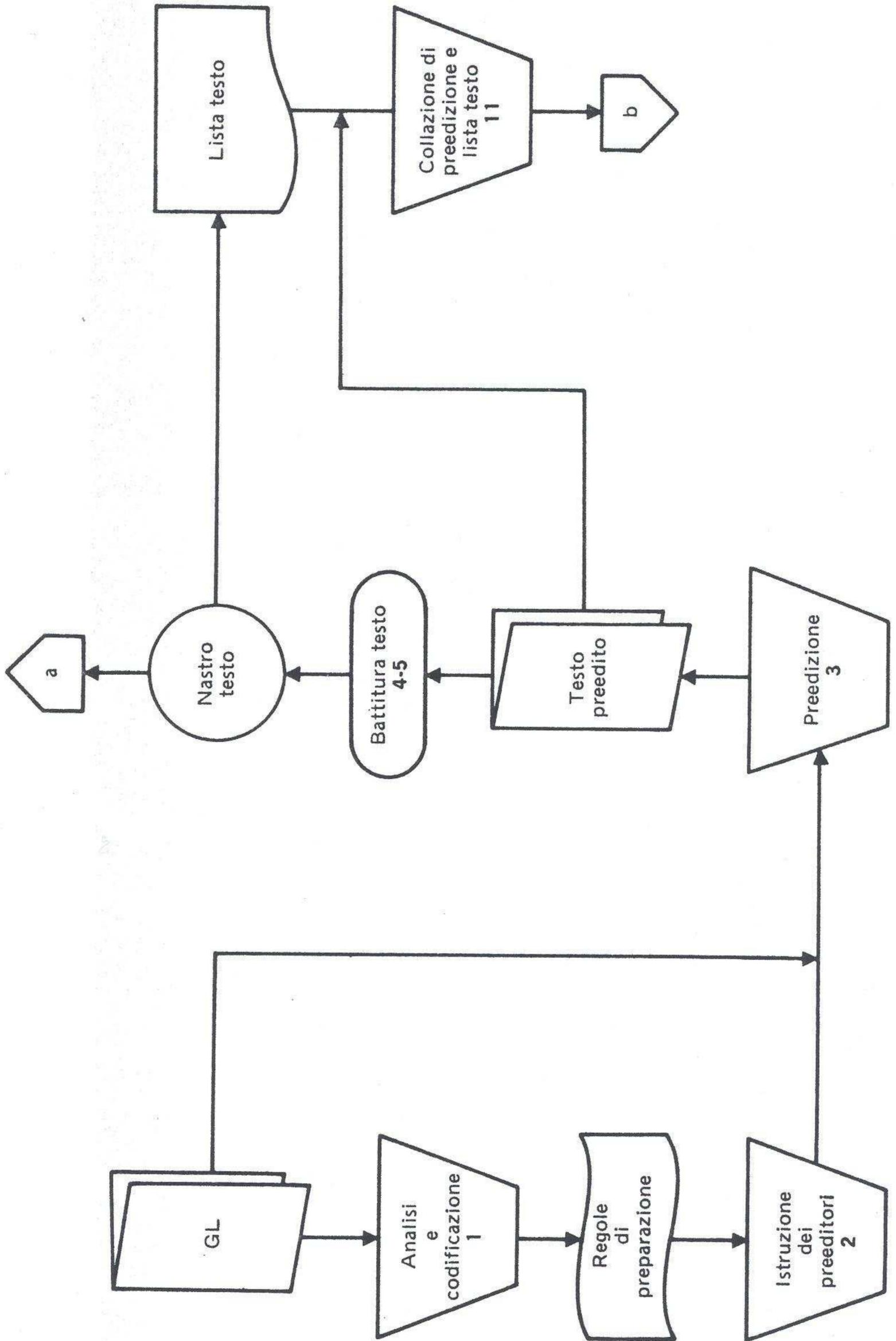
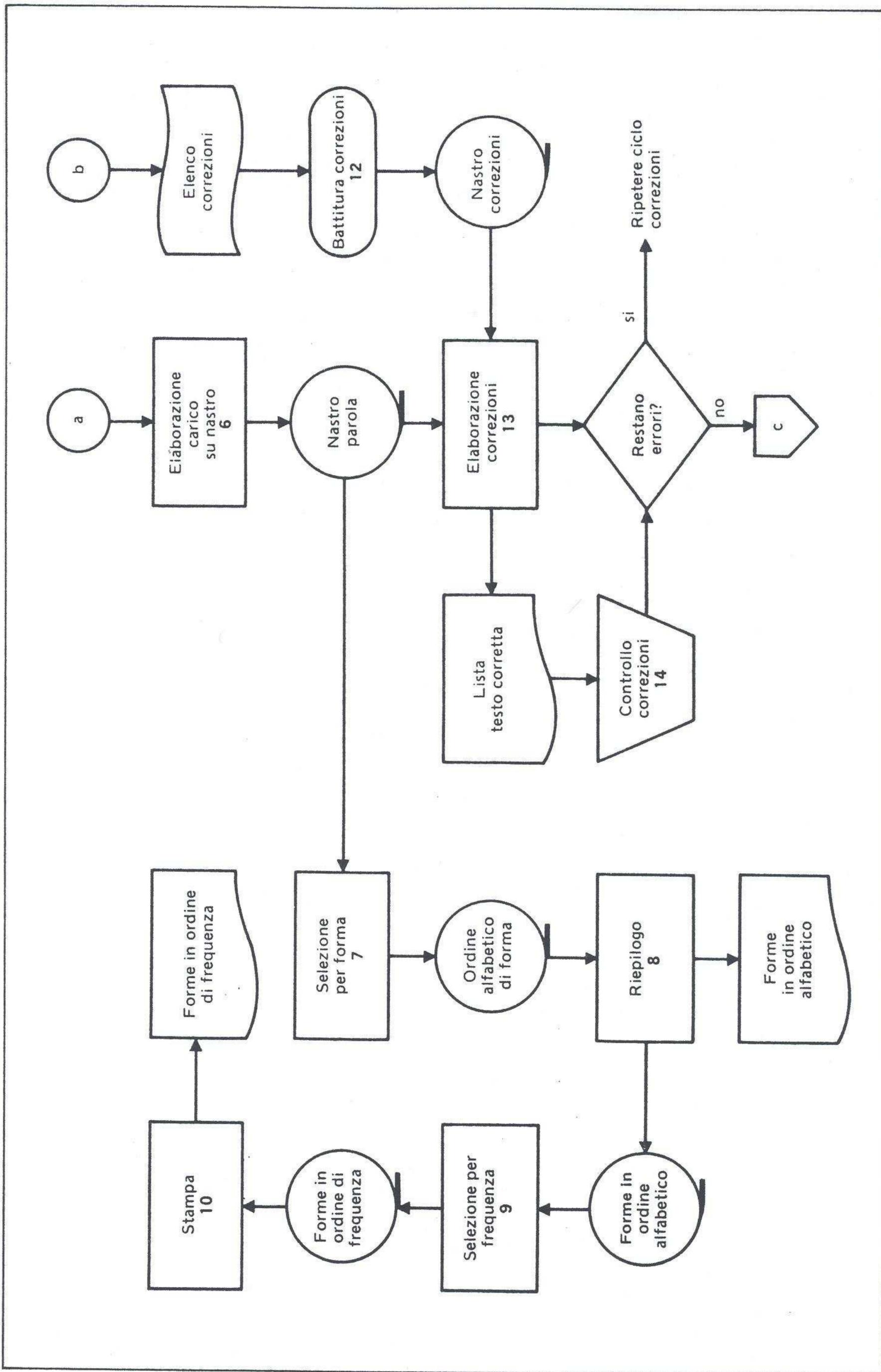
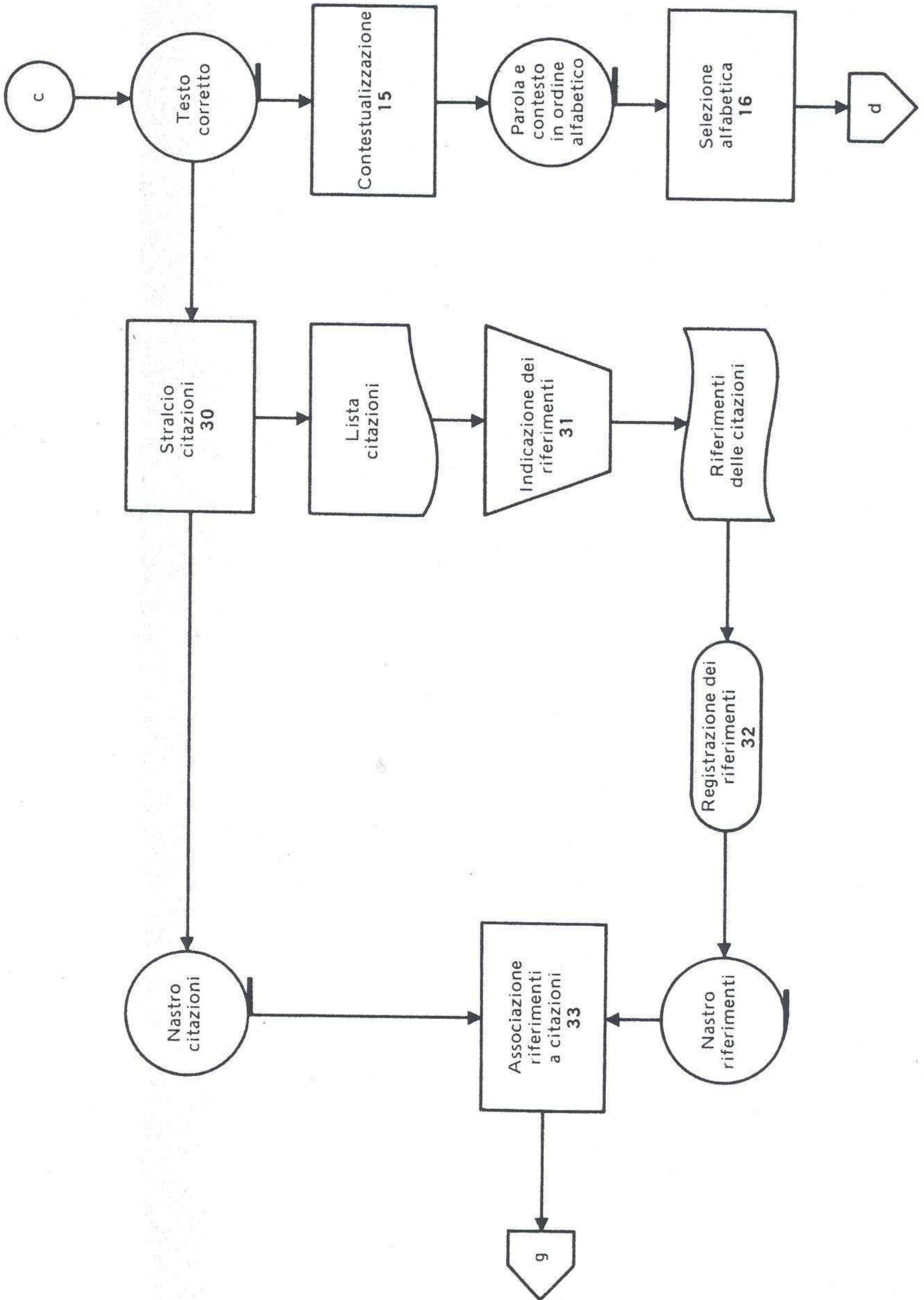
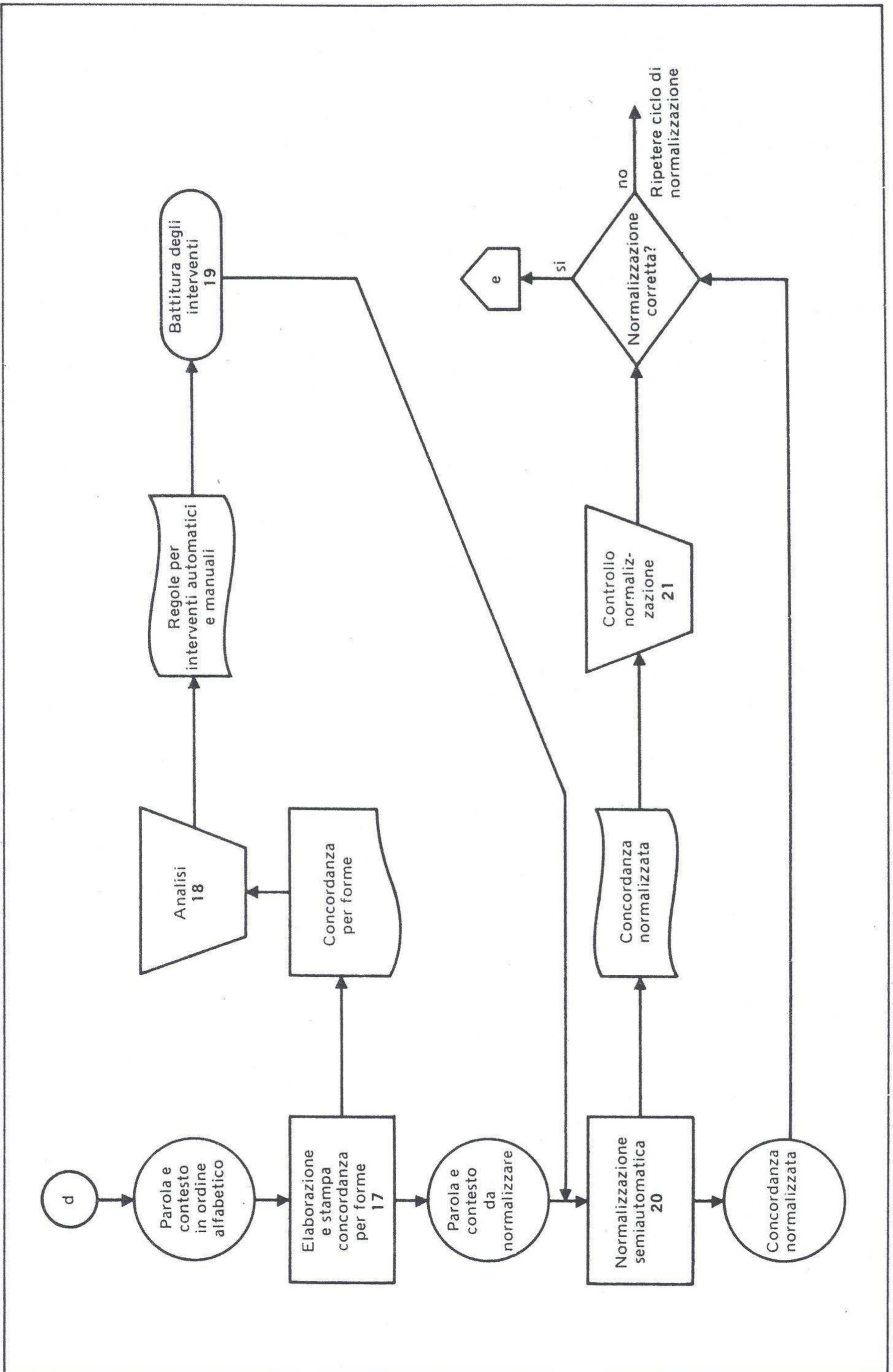


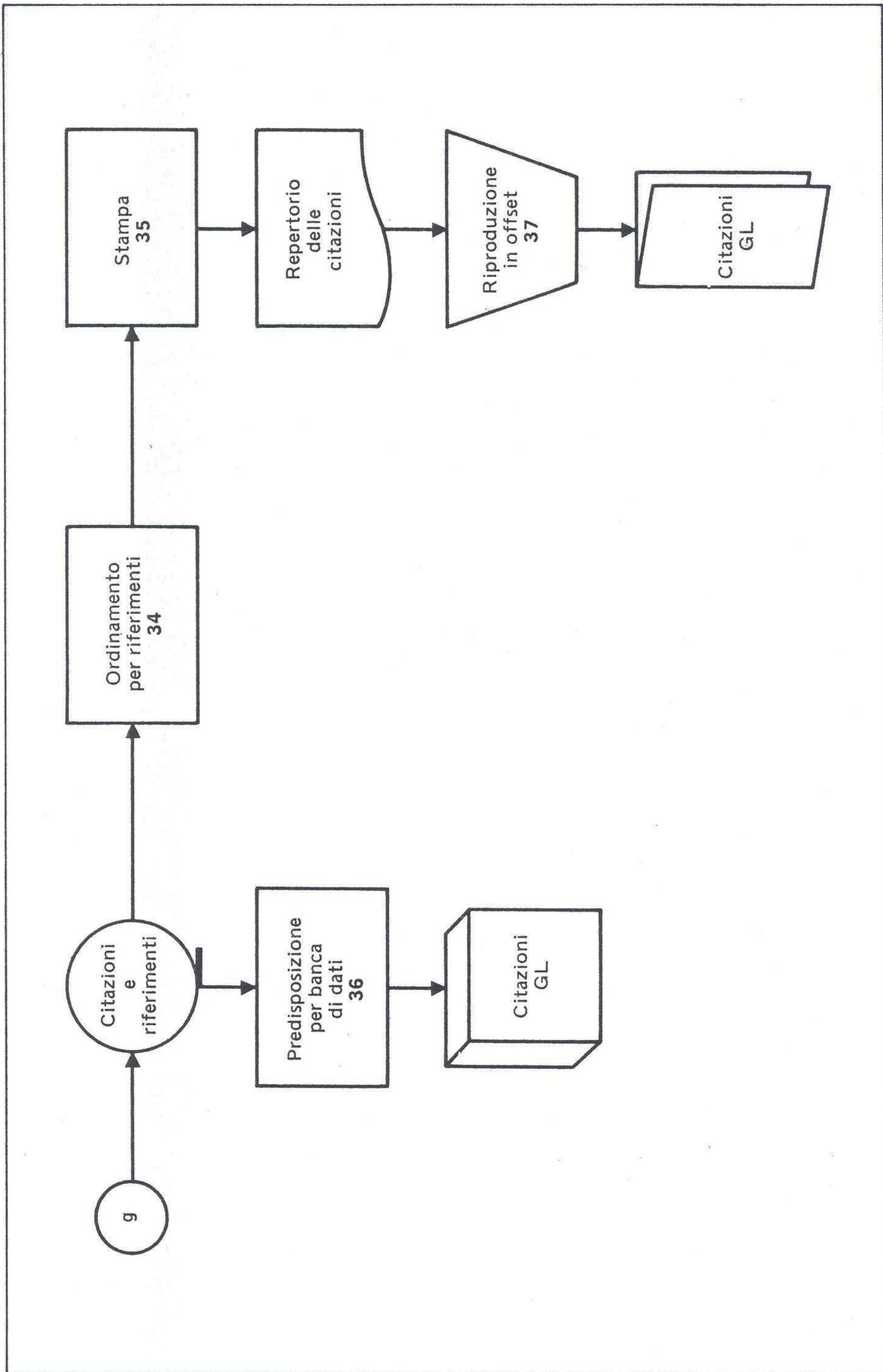
DIAGRAMMA OPERATIVO

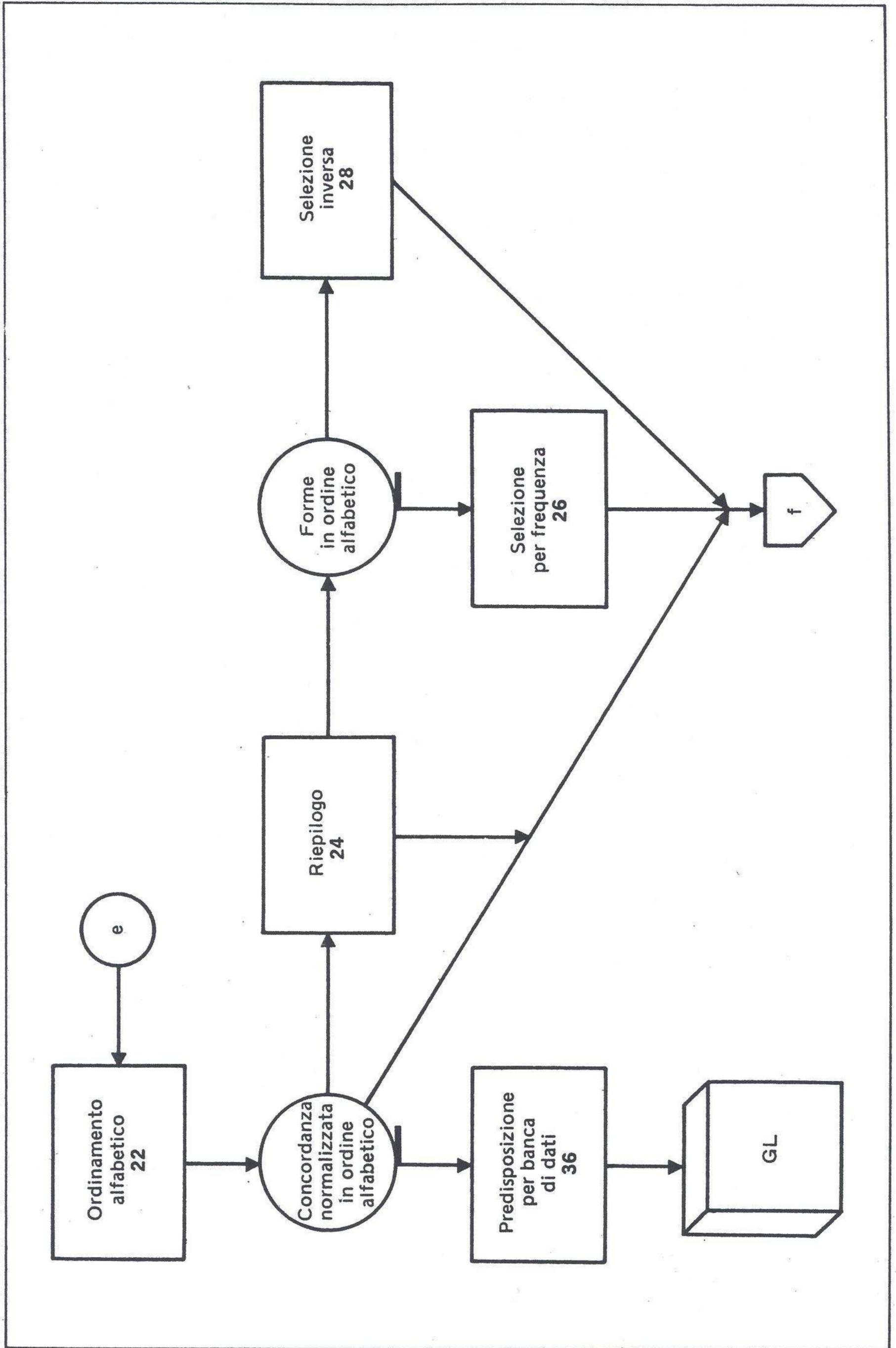


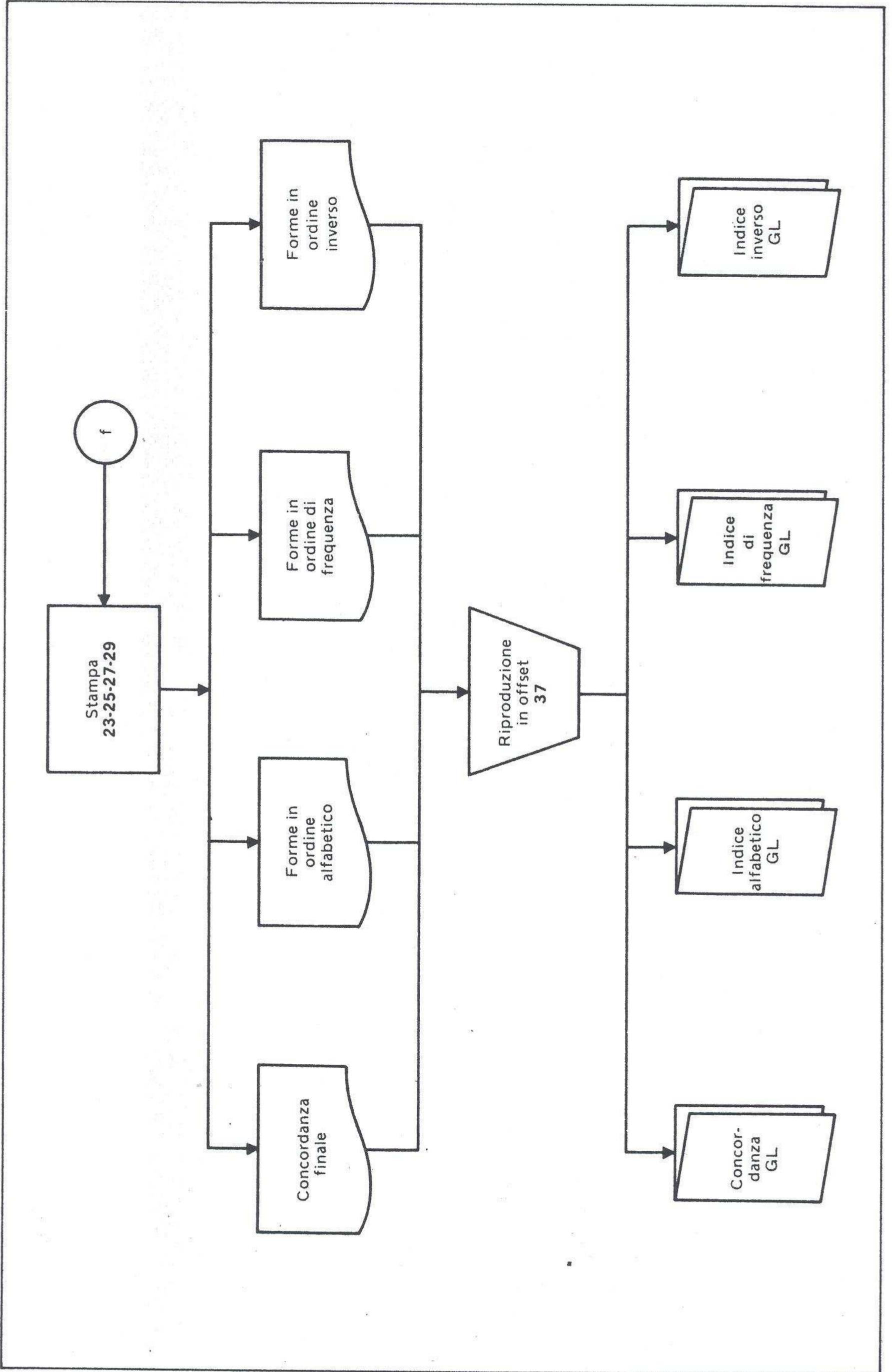












DESCRIZIONE DELLE OPERAZIONI.

op. 1. In una serie di riunioni, filologi e linguisti computazionali scelgono quali, tra i risultati possibili, convenga richiedere nel caso dei grammatici latini (6). In base all'analisi dei risultati da produrre e delle caratteristiche linguistico-filologiche del testo, vengono definiti i criteri di preedizione e le modalità di applicazione dello standard di codificazione e del sistema di spoglio del LLC.

op. 2. Vengono istruiti i collaboratori incaricati della preedizione e dei controlli successivi alla battitura.

op. 3. I collaboratori eseguono la preedizione: i testi vengono corredati con le informazioni elencate alle pp. 21 - 25 della nota di V. Lomanto.

op. 4. Come è noto, per ottenere che il testo sia leggibile dal calcolatore, è necessario ricopiarlo integralmente sulla tastiera di una macchina capace di tradurre i caratteri in fori su schede meccanografiche o in registrazione magnetica su nastro. Nel caso dei grammatici, abbiamo scelto questa seconda soluzione. Ogni riga battuta dall'operatore veniva immediatamente registrata su nastro magnetico.

La codificazione delle informazioni (riferimento, parole, punteggiatura, diverse modalità grafiche, distinzioni e aggiunte inserite in fase di preedizione), è avvenuta, come si è detto, secondo quanto prescritto dallo standard del LLC.

op. 5. Un operatore, diverso dal precedente, ribatte una seconda volta il testo sulla tastiera della stessa macchina, riga per riga. Questa volta, però, la riga appena battuta non viene registrata su nastro, ma viene confrontata, mediante un opportuno dispositivo automatico di cui la macchina è dotata, con la riga corrispondente memorizzata su nastro con la prima battitura. Se vi è una discordanza tra le due battiture, la macchina si blocca e l'operatore controlla se la discordanza è dovuta a un proprio errore o a un errore dell'operatore che lo ha preceduto. Nel secondo caso, la riga corretta va, ovviamente, a sostituire quella errata sul nastro.

op. 6. Il calcolatore 'legge' questo nastro, scompone il testo nelle diverse unità di elaborazione (nel nostro caso, le parole definite come sequenze di lettere tra due spazi o segni di interpunzione; in altre ricerche i grafemi, le sillabe, i sintagmi, ecc.) e lo registra parola per parola su un altro nastro magnetico, detto *nastro-parola*. Ogni parola costituisce una *unità indipendente* di registrazione record e riceve un numero progressivo che la individua univocamente.

Contemporaneamente, il calcolatore stampa il testo registrato, riproducendo il più fedelmente ed esplicitamente possibile tutte le informazioni registrate (*lista-testo*).

(6) Si vedano a questo proposito le note precedenti di Grilli e Marinone.

op. 7. Le parole del nastro-testo vengono ricopiate in ordine alfabetico su un altro nastro.

op. 8. Dal nastro delle parole ordinate alfabeticamente si ottiene un nuovo nastro con l'elenco alfabetico delle forme grafiche e delle rispettive frequenze. Questo elenco viene anche stampato.

op. 9. Le forme vengono disposte secondo l'ordine decrescente delle rispettive frequenze.

op. 10. Si stampa l'elenco delle forme in ordine di frequenza decrescente.

op. 11. Facendo la media su grandi quantità di righe perforate (qualche milione) si è constatato che, dopo la prima perforazione, le schede-testo contengono circa il 4% di schede errate, le quali con l'operazione di verifica vengono ridotte di solito allo 0,4%. Per scoprire questi ultimi errori, si esaminano gli elenchi alfabetici delle forme e si collaziona il testo stampato dal calcolatore con il testo originale. Gli errori trovati vengono elencati in moduli appositi, nei quali si riportano il numero progressivo che individua le registrazioni errate e le eventuali modifiche da introdurre.

op. 12. Dal modulo, le correzioni vengono ricopiate su un apposito nastro-correzione.

op. 13. Il calcolatore ricopia il nastro-parola correggendo gli errori secondo le indicazioni del nastro-correzione e contemporaneamente stampa una nuova lista-testo.

op. 14. Su questa lista-testo corretta, si controlla che gli errori identificati con l'op. 11 siano stati adeguatamente corretti. Il ciclo delle operazioni 11 - 14 viene ripetuto fino a quando non si trova più alcun errore.

op. 15. Il calcolatore ricopia le parole del testo corretto aggiungendo a ciascuna il relativo contesto.

op. 16. Le parole, con i relativi contesti, vengono ordinate alfabeticamente.

op. 17. Il calcolatore stampa la lista delle concordanze per forma; contemporaneamente ricopia su un nastro forme e relativi contesti numerati progressivamente, e con altri accorgimenti tecnici destinati a facilitare il ciclo di 'normalizzazione'.

A questo punto, di solito, si innesta la fase di lemmatizzazione, oppure, se la lemmatizzazione non è richiesta, lo spoglio viene considerato concluso.

La fase di lemmatizzazione ha di solito lo scopo di riunire sotto un unico esponente tutte le occorrenze di uno stesso lemma, le quali per lo più, nella concordanza delle forme grafiche, si trovano disperse sotto forme diverse.

Questa operazione comporta due momenti che, pur non distinti, di solito, dal punto di vista operativo, sono però distinti in linea di principio:

- a) *le forme grafiche vengono ricondotte alle rispettive forme lessicali*, e cioè:
 - a1) vengono ridotte differenze puramente grafiche (del tipo *adfero/affero*)
 - a2) vengono scomposte eventuali forme grafiche cui corrispondano due o più unità lessicali

(come, in italiano, le preposizioni articolate *o*, in latino, le forme con enclitica)

a3): vengono trattate in modo uniforme eventuali forme grafiche duplici (come italiano *sopra tutto/soprattutto*, o in latino *quando quidem/quandoquidem*)

a4) vengono raggruppate sotto due forme lessicali distinte le occorrenze di forme omografe.

b) le forme lessicali così ottenute e le loro occorrenze vengono assegnate ai lemmi di appartenenza.

Inoltre, nel caso di spogli selettivi, si utilizza di solito la fase di lemmatizzazione per operare la cernita dei lemmi e/o delle occorrenze da conservare.

Nel caso dei grammatici, pur essendo stata presa la decisione di non compiere la lemmatizzazione, si pensa di operare una 'normalizzazione' che ha lo scopo di facilitare il reperimento delle forme desiderate.

Questa normalizzazione corrisponde, sostanzialmente, alla somma degli interventi indicati ai punti *a1*, *a2*, *a3*.

Contemporaneamente, verranno contrassegnate le forme che devono essere eliminate dalle concordanze definitive, nonché eventuali categorie di forme che devono essere elencate in indici di 'cose notevoli'.

Poiché non verranno compiuti né gli interventi di tipo *a4* né quelli di tipo *b*, non parliamo di lemmatizzazione, ma solo di 'normalizzazione'. Mentre la scelta delle parole da espungere non può essere automatizzata, gli interventi di tipo *a1*, *a2*, *a3* possono essere automatizzati almeno in parte, nella misura in cui è possibile comunicare al calcolatore delle regole formali, traducibili in algoritmo, del tipo: «la stringa di caratteri φ deve essere trasformata nella stringa di caratteri ψ , nel contesto χ , nella condizione ω » (7).

È facile prevedere che regole di questo tipo, fondate esclusivamente su 'conoscenze' di carattere grafemico, risultino troppo 'potenti', nel senso che producono la trasformazione anche in casi nei quali invece essa non dovrebbe avere luogo. Tuttavia è conveniente, dal punto di vista operativo, applicarle ugualmente, sempre che il numero dei casi cui la regola si applica sia superiore a una certa soglia, e la regola tratti correttamente più della metà di questi casi. Basta infatti applicare la regola una prima volta con intenti puramente sperimentali ed euristici, chiedendo poi che il calcolatore stampi gli elenchi di tutte le forme trattate, ordinate opportunamente in modo da accelerare la consultazione. Prima di riapplicare la stessa regola, questa volta in modo definitivo, si contrassegnano le forme che, pur adempiendo alle condizioni formali richieste per l'applicazione della regola in questione, devono invece, di fatto, subire la trasformazione che essa specifica.

Può darsi che l'esame dei risultati suggerisca un perfezionamento delle regole, e che alcune forme costituiscano eccezioni che possono essere risolte solo comunicando direttamente al calcolatore, caso per caso, il loro trattamento. Per questo, è necessario prevedere una procedura di normalizzazione semiautomatica interattiva; nella quale, cioè, gli interventi umani e l'applicazione degli algoritmi si succedano ciclicamente, integrandosi a vicenda.

op. 18. Le concordanze vengono esaminate al fine di:

a) contrassegnare le forme da espungere nella stampa delle concordanze definitive (op. 23.);

b) contrassegnare le forme da includere in indici di 'cose notevoli';

(7) Per esempio, la stringa *dn* diventa *nn* nei contesti *a...i*, *a...o*, *a...u*, in posizione iniziale.

c) formulare le regole di normalizzazione, e specificare gli interventi da compiere per le eccezioni eventualmente già identificate all'atto stesso della formulazione delle regole. Le regole devono essere tradotte in algoritmi che vengono incorporati nel programma di 'normalizzazione'.

op. 19. I contrassegni (v. op. 18, *a* e *b*) e gli interventi (18 *c*) vengono memorizzati su nastro magnetico.

op. 20. Il programma di 'normalizzazione semiautomatica' viene applicato al nastro delle concordanze per forma. Esso ricopia su un nuovo nastro e contemporaneamente stampa le concordanze 'normalizzate', cioè concordanze nelle quali le forme-esponente hanno ricevuto i contrassegni e le modifiche specificate con l'op. 19, e sono state trasformate secondo le regole formulate con l'op. 18.

op. 21. Le concordanze 'normalizzate' vengono esaminate per controllare i risultati prodotti dalla op. 20. Se si riscontrano degli errori si ripete il ciclo delle operazioni 16 - 21 fino ad ottenere delle concordanze corrette.

op. 22. Le concordanze normalizzate e corrette vengono riordinate secondo l'ordine alfabetico delle forme-esponente: l'operazione di normalizzazione infatti ha tra i suoi scopi quello di creare le condizioni 'necessarie' per un diverso più perspicuo ordinamento.

op. 23. Si stampano le concordanze definitive, nelle quali non figurano le forme di cui è stata chiesta la espunzione. Di seguito, si stampano gli indici delle 'cose notevoli'.

op. 24 - 27. Si producono gli indici alfabetici e per frequenze decrescenti delle forme normalizzate, ripetendo le operazioni 7 - 10.

op. 28. Le forme vengono ordinate secondo l'ordine alfabetico inverso.

op. 29. Si stampa l'indice inverso delle forme normalizzate.

op. 30. Dal nastro-testo vengono ricopiate, su un nastro-stralcio, tutte le parti di testo che in fase di preedizione sono state racchiuse tra contrassegni che le qualificano come frasi o sintagmi che sono citazioni di altri autori. Contemporaneamente, viene stampata la lista di queste citazioni.

op. 31. Un filologo trascrive, accanto ad ogni citazione, il relativo riferimento (autore, opera, luogo), ovviamente quando sia noto.

op. 32. I riferimenti vengono ricopiati su nastro magnetico.

op. 33. Le citazioni vengono trascritte su un nuovo nastro accompagnate ciascuna dal proprio riferimento.

op. 34. Le citazioni vengono ordinate per autore, opera, passo.

op. 35. Si stampa l'indice degli autori e delle citazioni.

op. 36. I materiali prodotti dallo spoglio in 'machine readable form' (testo, forme 'normalizzate' con frequenze, contesti, e contrassegni speciali) confluiscono

nella banca dati, previa opportuna organizzazione che ne favorisca la consultazione e la manutenzione.

op. 37. I risultati prodotti dallo spoglio, dei quali si desidera la pubblicazione (concordanze selettive normalizzate, indici delle cose notevoli, indice delle citazioni, indici diretto, inverso, e per frequenza decrescente delle forme) vengono riprodotti in offset.

ANTONIO ZAMPOLLI
DOMENICO BROGNA