

DIPARTIMENTO DI STUDI UMANISTICI
DIPARTIMENTO DI SCIENZE GIURIDICHE, DEL LINGUAGGIO,
DELL'INTERPRETAZIONE E DELLA TRADUZIONE

Luciano Rocchi

Lessicologia
La classificazione delle lingue

LESSICOLOGIA

Lessico e vocabolario

La lessicologia è lo studio scientifico del **lessico**, l'insieme delle parole per mezzo delle quali i membri di una comunità linguistica comunicano tra loro (lessico italiano, tedesco, inglese). Il **vocabolario** (insieme di vocaboli) non è sinonimo di lessico, è invece un settore particolare, determinato del lessico. Tutte le parole usate ad esempio da un autore o da una scienza formano un vocabolario: il vocabolario del Manzoni (tutte le parole delle sue opere), il vocabolario della medicina (tutte le parole che riguardano la medicina).

L'unità di base del lessico è il **lessema**. I lessemi non sono soltanto quelle che comunemente si definiscono 'parole', ma anche i sintagmi che costituiscono un'unità lessicale (p.es. *per lo più, dopo cena*). Alcuni linguisti (non tutti) usano il termine 'lessema' anche come sinonimo di morfema lessicale (o radice).

Lessicalizzazione e grammaticalizzazione

Il lessico di ogni lingua è una parte che ha caratteristiche molto diverse da quelle proprie della grammatica, cioè si differenzia da fonetica, morfologia, sintassi. Mentre le strutture fonologiche, morfologiche e sintattiche sono insiemi stabili, non modificabili o aumentabili, il lessico invece è aumentabile e modificabile. Le strutture fonologiche, morfologiche e sintattiche sono sistemi chiusi, il lessico è un sistema aperto, suscettibile di variazioni e di arricchimenti. C'è un'opposizione tra fonemi e morfemi (di cui si può fare un inventario) da un lato e lessemi (di cui non si può fare un inventario) dall'altro, i primi costituiscono un repertorio chiuso, gli altri un repertorio aperto: abbiamo parole che escono dal lessico perché antiquate, obsolete o perché il loro referente è caduto in disuso o per motivi di moda, e d'altro canto abbiamo neologismi, parole che entrano nel lessico e lo modificano.

Questa opposizione però non è assoluta. Possono esserci morfemi che diventano lessemi: questo processo si chiama **lessicalizzazione**, o lessemi che diventano morfemi: in questo caso si parla di **grammaticalizzazione**. Parecchie preposizioni sono sorte da fenomeni di grammaticalizzazione: p. es. *durante* era orig. il participio presente di *durare* (*durante l'estate* = '(mentre) l'estate (è) durante'); *tranne* è l'imperativo cristallizzato di *trarre* + la particella pronominale *ne*: *tutti tranne* (= 'tranne, toglie') *me*; il franc. *chez* 'da' (*je vais chez le médecin* 'vado dal medico') deriva dal lat. *casa*, e in effetti in ant. franc. la parola significava 'casa', nel corso del tempo però tale significato è andato perduto (a favore di *maison*) e *chez* è sopravvissuto solo come morfema grammaticale: '(a) casa (di)' → 'presso, da'.

Per quanto riguarda la lessicalizzazione, citiamo casi come l'it. *il pro e il contro, il suo ex* (cioè ex-marito o ex-fidanzato), dove le preposizioni *pro*, *contro*, *ex* sono nominalizzate e acquistano quindi lo status di lessemi. Anche una sequenza di morfemi (un sintagma) può subire un processo di lessicalizzazione. A questo proposito, la formazione degli avverbi italiani in *-mente* fornisce un bell'esempio di entrambi i fenomeni sopra descritti. Quando il sintagma lat. *lenta mente* orig. 'con mente lenta' cominciò ad adoperarsi, con facile slittamento semantico, nel senso di 'in modo lento', *mente* non fu più percepito come ablativo di *mens*, ma come un semplice morfo di valore avverbiale, attraverso un processo di grammaticalizzazione. Nel contempo i due elementi del sintagma latino si fusero insieme (anche graficamente), arricchendo il lessico italiano di un nuovo lessema: *lenta-mente*. Alla grammaticalizzazione si è accompagnata dunque la lessicalizzazione.

Varietà del lessico

Il lessico non è una massa caotica, informe di parole, si può classificare e ordinare secondo criteri di diverso genere. Uno di questi consiste nel distinguere particolari settori del lessico in base all'uso che ne fanno i parlanti. Le principali varietà d'uso sono le seguenti:

1) **Varietà funzionali-contestuali**: sono i cosiddetti **linguaggi settoriali** (il linguaggio politico, pubblicitario, sportivo, tecnico-scientifico ecc.), quella porzione di lessico che riguarda ambiti particolari e specialistici. La differenza fondamentale tra linguaggio settoriale e lingua comune è che il linguaggio settoriale possiede vocaboli ed espressioni estranee alla lingua comune, o anche gli stessi vocaboli della lingua comune, usati però con significato diverso e specifico (v. oltre).

I linguaggi settoriali sono quella parte di lessico che è più sottoposta a rinnovamento (progresso scientifico, ecc.). Soprattutto il linguaggio tecnico-scientifico è ricco di neoformazioni che possono essere dei prestiti linguistici (dall'inglese), oppure composti che attingono dal latino e dal greco.

I composti di base greca e latina sono creazioni moderne modificanti l'aspetto originario delle parole classiche, adattandole agli scopi della lingua d'oggi: l'*anemometro* (misuratore del vento) viene da *ánemos* (vento) e *métron* (misura).

Ci sono anche composti formati da una parola greca più una parola latina: *aeronave* (greco + latino), *spettroscopio* (latino + greco).

Esistono dei composti con un elemento della lingua moderna e uno delle lingue classiche: *burocrazia* (francese + greco), *filmoteca* (inglese + greco).

Nelle varietà funzionali-contestuali si usano neoformazioni con suffissi e prefissi usati solo in questi linguaggi, ad esempio nel linguaggio della medicina il suffisso *-osi* che indica affezione cronica (artrosi, dermatosi, cirrosi), oppure *-ite* che indica affezione acuta (artrite, nefrite, gastrite).

Parole già esistenti nel lessico comune possono caricarsi di nuovi significati: p. es. il lessema italiano *campo* viene usato con accezioni del tutto particolari in molti linguaggi settoriali, come quelli della fisica (*campo magnetico*), del cinema (*campo lungo*), della linguistica (*campo semantico*), ecc. Il verbo *navigare*, che fino al Novecento significava unicamente 'andare per mare', grazie a una estensione semantica ha potuto essere accolto sia nel linguaggio dell'astronautica (*navigare attraverso lo spazio*) sia in quello informatico (*navigare in Internet*).

2) **Varietà geografiche** (definite anche **diatopiche**): una lingua parlata in un'area notevolmente estesa può presentare differenze tra un punto e l'altro della zona occupata dai suoi parlanti. L'italiano è un chiaro esempio di questo fenomeno: la lingua parlata nelle regioni italiane è diversa, non solo a livello di fonetica e morfologia, ma anche lessicalmente. Ci sono **regionalismi** lessicali che possono a loro volta suddividersi in dialettalismi. Nell'Italia settentrionale si dice *anguria* il frutto che in Toscana è chiamato *cocomero*. Questi termini che indicano lo stesso referente sono detti **geosinonimi**. C'è una distinzione tra regionalismi che restano tali e regionalismi che entrano nella lingua standard, anche se il confine tra gli uni e gli altri non è sempre facile da stabilire; in Italia, grazie soprattutto ai mezzi di comunicazione di massa, moltissimi regionalismi hanno trovato ampia diffusione e sono oggi patrimonio della lingua comune. Qualche esempio di antichi regionalismi ormai 'italianizzati': *lavandino* (che s'afferma sul tosc. *acquaio*), *metronotte*, *risotto*, *sberla* (dal lombardo); *abbaino*, *acciuga*, *mugugno* (dal ligure); *cicchetto*, *grissino* (dal piemontese); *ciao*, *giocattolo* (fino all'Ottocento si usava *balocco*, oggi ristretto al toscano), *vestaglia* (dal veneto); *abbacchio*, *fattaccio*, *imbonitore*, *intrufolarsi* (dal romano); *cafone*, *mozzarella*, *pizza*, *scocciare* (dal napoletano). Ancora forte connotazione regionale hanno invece p. es. il *ghisa* di Milano e il *pizzardone* di Roma ('vigile urbano'), nonché la nutrita serie di geosinonimi per 'ragazzo': *toso*, *bocia* (Veneto), *mulo* (Trieste), *pischello* (Roma), *guaglione*, *scugnizzo* (Napoli), *picciotto* (Sicilia).

3) **Varietà sociali** (definite anche **diastratiche**): nel lessico, nell'uso lessicale, si riflettono spesso le condizioni sociali di un individuo. La branca della linguistica che si occupa di questa problematica è la *sociolinguistica*. Le varietà sociali possono dipendere da diversi fattori, tra i quali citiamo:

- a) **L'età**: la parlata dei giovani è differente da quella degli anziani (gergo giovanile: *paninaro*, *fricchettone*, uso enfatico di aggettivi: *bestiale*, *mostruoso*). Questi termini sono generalmente soggetti a rapida usura. Chi si ricorda di *matusa*, che ebbe una stagione di grande (e altrettanto fugace) fortuna negli anni Sessanta?
- b) **Il sesso**: un'altra differenza sociale dal punto di vista del lessico riguarda i termini che vengono usati dagli uomini e non dalle donne (e viceversa); p. es. per quanto riguarda il vocabolario popolare inerente alla sfera sessuale le donne risultano più restie al suo utilizzo (o perlomeno così era un tempo). In lingue extraeuropee (soprattutto asiatiche) questa differenza è esplicitata in maniera macroscopica: in giapponese esiste un linguaggio maschile e uno femminile, che si diversificano per pronuncia, struttura morfosintattica e scelte lessicali.
- c) **Il livello di istruzione**: è ovvio che una persona (semi)analfabeta o che ha frequentato solo la scuola elementare avrà un vocabolario molto più limitato di chi ha proseguito gli studi e possiede una buona cultura.

La formazione delle parole

Come s'è già accennato, il lessico è la parte di un sistema linguistico maggiormente esposta ai mutamenti, una realtà in continua evoluzione. I procedimenti attraverso i quali si arricchisce il lessico, introducendo nuovi lessemi, sono essenzialmente due: la **formazione delle parole** e il **prestito** (di quest'ultimo ci occuperemo più tardi).

La formazione delle parole riguarda quel complesso di trasformazioni attraverso cui si passa da una parola base già esistente nel lessico ad una parola non esistente prima, nuova, che si definisce **neoformazione**. Questi processi di formazione sono divisibili in due categorie: **derivazione** e **composizione**. La derivazione, che naturalmente è possibile soltanto in lingue flessive o agglutinanti (cioè con struttura morfologica che la consenta), consiste nell'aggiunta di materiale morfologico ad una parola, ossia nell'aggiunta di affissi, che si suddividono in suffissi, prefissi e infissi, a seconda che seguano, precedano o siano inseriti nella parola base. Tralasciando gli infissi, che hanno un uso piuttosto limitato, parleremo brevemente dei processi di suffissazione e prefissazione.

Suffissazione e prefissazione

La **suffissazione** è caratterizzata dal fatto di poter contemplare il passaggio da una categoria di parole a un'altra, cioè un verbo può dar luogo a un nome o a un aggettivo, un nome a un verbo o a un aggettivo, un aggettivo a un verbo o a un nome, ecc. Le neoformazioni che derivano da un nome si chiamano **denominali**, da un aggettivo **deaggettivali**, da un verbo **deverbali**. Diamo qualche esempio italiano:

DENOMINALI	<i>romanzo</i> → <i>romanziere</i>	nominale
	<i>romanzo</i> → <i>romanzesco</i>	aggettivale
	<i>romanzo</i> → <i>romanzare</i>	verbale
DEAGGETTIVALI	<i>bianco</i> → <i>bianchezza</i>	nominale
	<i>bianco</i> → <i>biancastro</i>	aggettivale
	<i>bianco</i> → <i>biancheggiare</i>	verbale
DEVERBALI	<i>lavorare</i> → <i>lavorazione</i>	nominale
	<i>lavorare</i> → <i>lavorabile</i>	aggettivale
	<i>lavorare</i> → <i>lavoricchiare</i>	verbale

Esistono anche **deverbali a suffisso zero**: p. es. *comando* da comandare, *delibera* da deliberare, *sbaglio* da sbagliare; sono nominali tratti da verbi, privi però di particolari suffissi.

Vanno ancora rammentati i numerosissimi verbi **parasintetici**, in cui si riscontra l'utilizzo simultaneo di un suffisso e un prefisso: *abbottonare* da bottone, *imbottigliare* da bottiglia, *sbandierare* da bandiera, ecc.

Una particolare forma di suffissazione è l'**alterazione**, attraverso la quale il significato della parola base non muta nella sostanza, ma soltanto per alcuni particolari aspetti. Gli alterati possono essere diminutivi, vezzeggiativi, accrescitivi, peggiorativi (*omino*, *ometto*, *omarino*, *omaccione*; *casina*, *casetta*, *casuccia*, *casaccia*). L'italiano è una lingua particolarmente ricca di alterati grazie al numero piuttosto elevato di suffissi alterativi che esso possiede; in altre lingue invece l'alterazione si presenta molto più limitata, p. es. il nostro *cagnolino* in inglese e francese non può essere reso che con un sintagma del tipo 'piccolo cane': *small dog*, *petit chien*. Il tedesco ha unicamente due suffissi alterativi diminutivi, *-chen* e *-lein* (*Haus* 'casa' ~ *Häuschen* 'casetta', *Frau* 'signora' ~ *Fräulein* 'signorina').

La **prefissazione**, a differenza della suffissazione, non comporta il mutamento di categoria, il nome rimane nome, il verbo verbo, ecc. P. es. *dare* → *ridare*, *fedele* → *infedele*, *avviso* → *preavviso*. I prefissi sono di varia natura, per lo più nelle lingue europee erano originariamente preposizioni o avverbi (diamo alcuni esempi dell'italiano):

- *anti-*, *ante-* indicano anteriorità spaziale o temporale: *anticamera*, *anteguerra*
- *contro-*, *contra-* indicano opposizione: *contraccolpo*, *controsenso*
- *sopra-*, *sovra-* esprimono superiorità, in senso proprio e figurato: *sopravveste*, *sovrumano*
- *inter-* indica 'in mezzo': *internazionale*, *interregno*
- prefissi intensivi come *iper*, *super*: *ipersensibile*, *supermercato*
- prefissi negativi come *in-*, *s-*, *dis-*, *non-*: *incapace*, *scontento*, *disonore*, *nonsense*.

Composizione

La composizione consiste nella formazione di nuovi lessemi, chiamati **composti**, attraverso l'unione di lessemi preesistenti (mentre gli affissi che entrano in gioco nella derivazione sono morfemi grammaticali, non lessicali). Il fenomeno della composizione ha un'incidenza diversa da lingua a lingua. Tra le lingue europee, la privilegiano in modo speciale quelle germaniche e ugrofinniche, dove i composti raggiungono un numero illimitato; p. es. se prendiamo le parole per 'scuola' e 'compagno' in tedesco (*Schule*, *Kamerad*), inglese (*school*, *mate*), ungherese (*iskola*, *társ*) e finlandese (*koulu*, *toveri*) dalla loro unione risulterà agevolmente un nuovo lessema dal significato 'compagno di scuola': *Schulkamerad*, *schoolmate*, *iskolatárs*, *koulutoveri*. Ciò non è possibile nelle lingue romanze o in quelle slave, nelle quali la composizione agisce in maniera più limitata.

I composti si possono suddividere in due tipi fondamentali: **composti con base verbale** e **composti con base nominale**:

- 1) I composti con base verbale sono quelli che, se sciolti nei loro componenti, esprimono una frase con predicato verbale: *accendisigari* (indica 'qualcosa accende i sigari'), *aspirapolvere* ('qualcosa aspira la polvere'). In italiano l'inventario di questi composti è piuttosto consistente: *asciugamano*, *lavapiatti*, *scaldabagno*, ecc. Nei linguaggi settoriali troviamo numerosi composti formati con elementi di origine greca, dove l'ordine degli elementi è inverso rispetto all'uso italiano, con la base verbale al secondo posto: *antropofago*, *telegrafo*, *cardiopatia* (*-fago* 'mangiare', *-grafo* 'scrivere', *-patia* 'soffrire').
- 2) I composti con base nominale, formati cioè da due elementi nominali, si suddividono innanzitutto tra **subordinativi**, che presuppongono tra i loro componenti un rapporto di subordinazione, e **coordinativi**, che invece ne presuppongono uno di coordinazione. Al primo tipo appartengono p. es. *capostazione* ('capo della stazione') e *agopuntura* ('puntura con l'ago'). In italiano non sono molto frequenti, mentre sono comunissimi, come s'è detto sopra, nelle lingue germaniche. Per quanto concerne i composti coordinativi, nella nostra lingua sono formati soprattutto da no-

me + aggettivo (o aggettivo + nome) e sostituiscono frasi con un predicato nominale: *terraferma* ('la terra è ferma'), *cassaforte* ('la cassa è forte'), *camposanto*, *altopiano*, *malafede*, *bassorilievo*, ecc.

Un altro tipo di composto nominale, chiamato anche **composto possessivo**, presuppone un riferimento esterno diverso rispetto ai suoi costituenti: ingl. *redskin*, it. *pellerossa* = (X ha la) pelle rossa, ingl. *flatfoot*, it. *piedi piatti* = (X ha i) piedi piatti, lat. *magnanimus* = (X ha un) grande animo..

Infine abbiamo composti nominali che sostituiscono due predicati nominali coordinati: *cassapanca* ('qualcosa che è insieme cassa e panca'), *agrodolce* ('qualcosa che è insieme agro e dolce'), *sordomuto*. Composti di questo genere sono frequenti per indicare i colori che contraddistinguono le maglie delle squadre di calcio: *bianconero*, *rossonero*, *nerazzurro*.

Un tipo particolare di composizione è quella rappresentata dai cosiddetti **conglomerati**. Si tratta di segmenti dell'enunciato irrigiditisi nell'uso fino a diventare unità lessicali autonome: it. *nontiscordardimé*, ingl. *forget-me-not* 'nome pop. della miosotide', franc. *va-nu-pieds* ('va a piedi nudi') 'straccione, pezzente'. In italiano troviamo parecchi conglomerati formati da imperativi cristallizzati: *saliscendi*, *toccasana*, *fuggifuggi*, *dormiveglia*.

Prestito linguistico

Quando due sistemi linguistici diversi vengono a contatto e interagiscono si produce quella che viene chiamata **interferenza linguistica**. Il fenomeno più appariscente dell'interferenza è il **prestito** (termine fuorviante, dato che non è la cessione di qualcosa che poi viene restituito): esso consiste nell'utilizzazione da parte di una lingua B di un tratto linguistico che esisteva precedentemente in una lingua A e che non era posseduto da B. Si tratta in buona sostanza di un processo di carattere mimetico, cioè di imitazione: il parlante ha di fronte un modello straniero e si sforza di acquisirlo, cercando di imitarlo e di riprodurlo, di ricrearlo all'interno del proprio sistema linguistico.

Il prestito è un fenomeno antichissimo, che risale alla notte dei tempi; nessuna lingua, a meno che i suoi parlanti non vivano in completo isolamento, è immune dal prestito. Se prendiamo il lessico italiano, buona parte di esso è formato da prestiti, entrati nella nostra lingua in epoche diverse e per vie diverse. Lo strato più antico è rappresentato dai germanismi, assunti durante le invasioni barbariche da Goti, Longobardi e Franchi, p. es. *sapone*, *guerra*, *elmo*, *schietto*, *spalla*, *guancia*, *schiena*, *scherzare*, *spaccare*, *graffiare*. Dall'arabo, attraverso la Sicilia, ci sono arrivati, tra gli altri, *arancia*, *limone*, *carciofo*, *spinaci*, *dogana*, *magazzino*, *tariffa*. Con l'avvento della civiltà cortese abbiamo i francesismi: *cavaliere*, *dama*, *gioiello*, *cuscino*, *gonfalone*, ecc. Alla dominazione spagnola nel Cinquecento e Seicento sono dovuti *compleanno*, *complimento*, *puntiglio*, *sfarzo*, ecc. La lingua parlata in un'area adiacente a quella di un'altra lingua e che entra con essa in rapporto di reciproca influenza prende il nome di **adstrato** (p. es. i francesismi in italiano e gli italianismi in francese sono dovuti all'adstrato), mentre il complesso degli elementi linguistici giunti con una lingua importata dall'estero che per un certo tempo ha convissuto con la lingua indigena forma il **superstrato** (p. es. i germanismi sopra elencati fanno parte del superstrato germanico dell'italiano).

Anche molte parole italiane di origine latina si possono in un certo senso considerare prestiti, in quanto non provenienti direttamente dal latino parlato, ma recuperate per via dotta dal latino classico. Non è raro che la stessa parola latina abbia dato in italiano due esiti, l'uno popolare e l'altro dotto; tali doppioni si definiscono **allotropi**. P. es. *vitium* per via popolare si è evoluto in *vezzo*, mentre *vizio* è un latinismo dotto. Diciamo quindi che 'vezzo' è l'allotropo popolare di 'vizio' o, all'inverso, che 'vizio' è l'allotropo dotto di 'vezzo'. Altri esempi: da *circulus*: *cerchio* (popolare) e *circolo* (dotto), da *verecundia*: *vergogna* (popolare) e *verecondia* (dotto). Può capitare che anche una parola straniera dia luogo ad allotropi; citiamo il caso di *arsenale* e *darsena*: entrambi derivano dall'arabo *dār aṣ-ṣinā'a* (lett. 'casa di costruzione'), il primo però è giunto in Italia attraverso il veneziano, l'altro attraverso il genovese.

Tipi di prestito

In base alle motivazioni che spingono i parlanti a mutuare parole straniere, i prestiti si dividono in due tipi: **di necessità** e **di lusso**. Il primo si ha quando la parola acquisita col prestito indica un referente inesistente nella lingua di arrivo, p. es. *patata* (di origine amerindia, tramite lo spagnolo), pianta sconosciuta in Europa fino alla scoperta dell'America, *caffè* (dal turco), bevanda da noi del tutto ignota prima dei contatti col mondo ottomano, *zero* (dall'arabo), dato che nella numerazione romana non esisteva lo zero, *canguro*, animale australiano il cui nome viene da una lingua locale.

Il prestito di lusso ha invece fine stilistici o di promozione sociale, con esso il parlante tende a preferire forme considerate di maggior prestigio o più confacenti ai suoi bisogni espressivi. P. es: *leader*, *babysitter*, *chic*, *killer* sono in italiano prestiti di lusso perché potrebbero tranquillamente essere sostituiti da *capo*, *bambinaia*, *elegante*, *sicario*. A volte il loro uso è determinato da motivi di praticità (*flirt* è più economico di 'breve relazione amorosa'), più spesso da pigrizia o snobismo.

Integrazione e acclimatamento dei prestiti

Il processo attraverso il quale una lingua acquisisce un prestito non è, per così dire, indolore. Infatti le strutture fonologiche e morfologiche sono diverse per ciascuna lingua e capita perciò molto spesso che strutture proprie della lingua di partenza non siano compatibili con quelle della lingua d'arrivo. Pertanto è inevitabile che il prestito debba in qualche modo adeguarsi alle strutture della lingua che lo riceve. Questo adeguamento è chiamato **integrazione** e può essere di vari tipi:

- 1) **fonetica**: è l'integrazione più naturale, quella che si potrebbe definire indispensabile, in quanto da lingua a lingua varia non solo il repertorio fonemico, ma spesso anche la realizzazione di uno stesso fonema, a seconda della variante allofonica prescelta. Perciò un prestito, per quanto possa mantenere il suo aspetto straniero, sarà sempre integrato foneticamente, cioè adattato alle abitudini articolatorie della lingua d'arrivo. P. es. la pronuncia italiana corrente di *film*, *computer*, *routine*, cioè /film/, /kompju:ter/, /rutin/, è ben diversa da quella inglese /film/, /kəmˈpju:tə/ e francese /rutin/.
- 2) **morfologica**: è quel complesso di modificazioni a cui è sottoposto un prestito per renderlo compatibile con la struttura morfologica della lingua d'arrivo. A volte un prestito può adattarsi senza difficoltà alla morfologia della lingua che lo riceve: p.es. in tedesco *Reporter*, *Boxer*, anglicismi, vengono allineati senza sforzo ai nomi d'agente autoctoni come *Lehrer* 'insegnante'; il finl. *sauna* (che non ha genere) grazie alla sua terminazione in -a ha potuto inserirsi con tutta facilità nella categoria dei femminili italiani (la sauna, le saune). In italiano questo però è un caso raro, dato che la maggior parte delle lingue da cui abbiamo attinto (e continuiamo ad attingere) prestiti hanno una struttura morfologica (terminazioni in consonante) diversa dalla nostra. Possiamo dire che quasi tutti i prestiti entrati in italiano prima dell'Ottocento hanno subito un processo di integrazione morfologica: *elmo* (< got. *hilms*), *cavaliere* (< provenz. *cavalier*), *carciofo* (< ar. *ḥuršūf*), *lanzicheneco* (< ted. *Landsknecht*), *bistecca* (< ingl. *beefsteak*), ecc. Nel corso degli ultimi due secoli invece assistiamo a una progressiva non-integrazione dei forestierismi, di cui è rispettata la morfologia originaria: *bar*, *sport*, *film*, ecc. Se fossero integrati si direbbe **filmo*, **sporto*...

L'integrazione morfologica può riguardare anche aspetti diversi dalla desinenza. Nelle lingue germaniche i verbi forti, caratterizzati dall'alternanza apofonica, appartengono di norma al fondo lessicale originario. Un'eccezione notevole è costituita dal ted. *schreiben* (pret. *schrieb*, part. pass. *geschrieben*) 'scrivere', che è un latinismo; ciò si deve al fatto che l'antico alto ted. *scrīban*, prestito dal lat. *scrībere*, è stato assimilato, grazie al suo vocalismo radicale, ai verbi del tipo di *trīban* (mod. *treiben*, *trieb*, *getrieben*) 'spingere'. In swahili *kitabu* 'libro' è un prestito dall'ar. *kitāb*; i parlanti questa lingua hanno però reinterpretato la sillaba iniziale dell'arabismo identificandola col prefisso *ki-* che contraddistingue una delle loro classi nominali (p. es. *kitu*

‘cosa’, *kisu* ‘coltello’), col risultato che *kitabu* viene inserito perfettamente in un paradigma swahili: il suo plurale suona *vitabu*, esattamente come *vitu* e *visu* sono i plurali di *kitu* e *kisu*.

- 3) **grafica**: si ha quando, riproducendo foneticamente un modello straniero, si usano le norme grafematiche della lingua d’arrivo e non quelle della lingua di partenza, nel caso le prime divergano dalle seconde. P. es. i termini francesi *bureau* e *liqueur* sono diventati in tedesco *Büro* e *Likör*, ossia per la resa di /y/, /o/, /k/, /ø/ i grafemi tedeschi (ü, o, k, ö) hanno sostituito quelli francesi (u, eau, qu, eu). Parimenti i due lessemi che compongono il sintagma turco *otobüs şoförü* ‘conduttore d’autobus’ sono tolti di peso dal francese (*autobus* e *chauffeur*), ma risultano pienamente integrati nel sistema grafico turco. L’it. *sci* è un prestito dal norv. *ski*, che si pronuncia /ʃi/; quando la parola è stata introdotta in italiano, verso la fine dell’Ottocento, si usava la grafia originaria con la k, ma a partire dagli anni Venti è intervenuta l’integrazione grafica, si è cominciato cioè a scrivere *sci* per riprodurre l’effettiva pronuncia norvegese.

Talvolta uno stesso prestito si può riscontrare in forma sia integrata sia non integrata (senza specificazioni, ci si riferisce sempre all’integrazione morfologica). Basteranno un paio d’esempi: nell’area italiana, il francesismo *toilette* - non integrato a livello letterario - si mostra invece variamente integrato nell’uso popolare (*toiletta*, *toletta*, *teletta*) e lo stesso si può dire di *chic*, che nel toscano parlato diventa *scicche* (qui, oltre all’integrazione morfologica, si nota anche quella grafica, col digramma italiano *sc* a sostituire il francese *ch*).

Può accadere che manchi la riproduzione fonetica del modello straniero e che i suoi elementi grafematici vengano pertanto assimilati, nella pronuncia, a quelli della lingua d’arrivo. Per tornare a un esempio succitato, *ski* in inglese suona /ski:/: la mancata integrazione grafica del prestito norvegese ha fatto sì che il nesso *sk* fosse letto dai parlanti anglofoni come nelle parole della loro lingua *skin*, *skill*, ecc. *Tunnel* viene da noi pronunciato ‘all’italiana’, mentre un’imitazione del modello inglese /tʌnl/ avrebbe dovuto dare qualcosa come **tanel* o **tonel*. Il termine dialettale veneto *schei* ‘soldi’ deriva dalla prima sillaba, letta secondo le nostre consuetudini grafiche, del ted. *Scheidemünze* /ʃaidəmyntsə/ ‘moneta divisionale’, che appariva scritto sui centesimi all’epoca del Lombardo-Veneto.

Non bisogna confondere l’integrazione con l’**acclimatemento**, che riguarda il grado di familiarità, la frequenza d’uso di un prestito. E’ piuttosto comune il caso di parole non integrate come *sport*, *bar*, *stop*, *camion*, che sono però perfettamente acclimate, come dimostrano sia la generalizzazione del loro impiego, sia i loro vari derivati (*sportivo*, *barista*, *stappare*, *camioncino*).

Paretimologia

Nel processo di acquisizione di un prestito può intervenire, alterando più o meno profondamente la resa del modello straniero, la **paretimologia** (detta anche *etimologia popolare* o *falsa etimologia*). Si tratta di un complesso fenomeno attraverso il quale un tratto linguistico viene arbitrariamente modificato per renderlo trasparente, e quindi motivato, agli occhi del parlante. Alcune esemplificazioni permetteranno di chiarire il concetto. L’ingl. *crayfish* ‘gambero’ è un prestito dall’ant. franc. *crevice*; la terminazione *-vice* è stata modificata in *-fish* dai parlanti inglesi per accostamento paretimologico a *fish* ‘pesce’ e in base all’analogia coi nomi di altri animali acquatici quali *cuttlefish* ‘seppia’ e *jellyfish* ‘medusa’. Se dal ted. *Steinbock*, invece dell’atteso **stambocco*, abbiamo l’it. *stambecco*, ciò dipende dall’influsso paretimologico di *becco* ‘caprone’, a cui lo stambecco è facilmente avvicinabile. La resa ital. *stoccafisso* dell’ol. *stokvis* (lett. ‘pesce-bastone’) è certamente dovuta all’intromissione del nostro aggettivo *fisso*, con allusione alla rigidità di questo pesce. L’ingl. *country dance* ‘danza campestre’ è diventato in Francia *contredanse* (da cui l’it. *contraddanza*); i parlanti francesi hanno rimotivato la prima parte del sintagma inglese trasformandola nel per loro trasparente prefisso *contre-*. Infine citiamo il caso di lat. *asparagus*, che nell’ingl. dial. ha dato luogo a *sparrow*, confusi ben presto con una voce omofona significante ‘passero’ e ampliato di con-

sequenza con l'aggiunta di *grass* 'erba'; l'asparago è stato quindi paretimologicamente reinterpretato come *sparrowgrass* 'erba del passero'. Ci troviamo di fronte a un completo rimodellamento del prestito latino, che ha comportato una rimotivazione semantica.

Calco linguistico

Il calco consiste, come il prestito, nella ripresa di un modello straniero da parte di una lingua X; però mentre attraverso il prestito la lingua X si arricchisce di nuovi lessemi (che riproducono formalmente il modello), attraverso il calco tale modello viene riprodotto utilizzando materiale lessicale già esistente nella lingua X. Abbiamo due tipi fondamentali di calco: **calco strutturale** (chiamato anche *calco di traduzione*) e **calco semantico**.

Calco strutturale

Il calco strutturale si produce quando un composto o un sintagma viene trasposto in un'altra lingua, traducendo più o meno letteralmente gli elementi da cui è formato. Esso è divisibile in vari tipi:

- 1) **calco perfetto**: riproduce il modello straniero rispettandone fedelmente l'ordine degli elementi, cioè la struttura sintattica: *fuorilegge* e *ferrovia* sono calchi italiani perfetti dell'ingl. *outlaw* (fuori + legge) e del ted. *Eisenbahn* (ferro + via). Naturalmente i calchi perfetti sono frequenti soprattutto tra lingue che hanno strutture sintattiche simili: ingl. *summit conference* 'conferenza al vertice' → ted. *Gipfelkonferenz*, ingl. *brainwashing* 'lavaggio del cervello' → ted. *Gehirnwäsche* (inglese e tedesco condividono l'ordine regressivo). Allo stesso modo sintagmi italiani come *libero pensatore*, *messa in scena*, *colpo di stato* calcano perfettamente quelli francesi *libre penseur*, *mise en scène*, *coup d'état*.
- 2) **calco imperfetto**: la ripresa del modello straniero non è fedele o sintatticamente o lessicalmente. Il primo caso è comunissimo, quasi tutti i numerosi calchi italiani sull'inglese o sul tedesco sono imperfetti (le strutture sintagmatiche e compositive italiane sono di norma progressive, a differenza di quelle inglesi e tedesche): cfr. *skyscraper* → *grattacielo* (se il calco fosse perfetto avremmo **cielograttatore!*), *round-table* → *tavola rotonda*, *iron curtain* → *cortina di ferro*, e v. sopra le rese italiane di *summit conference* e *brainwashing*. Sul tedesco: *Arbeitgeber* → *datore di lavoro*, *Klassenkampf* → *lotta di classe*. La mancata corrispondenza sul piano lessicale è bene illustrata da un paio di esempi tedeschi: *Wolkenkratzer* 'grattacielo' (lett. 'grattanuvole'), dove allo *sky* del modello inglese si è sostituito *Wolke* 'nuvola'; *Halbinsel* 'penisola' (lett. 'mezza isola') rende solo approssimativamente il lat. *paeninsula*, il cui primo elemento *paene* significa 'quasi' (e difatti il franc. *presqu'île* ne è il calco perfetto).
- 3) **calco parziale**: uno solo degli elementi del modello è tradotto, mentre l'altro viene riprodotto con fedeltà, insomma la parola che ne risulta è per metà calco e per metà prestito. Qualche esempio: ingl. *tramway* → it. *tramvia* (*way* = via; l'altra resa *tranvai* rientra invece in quei casi di pronuncia 'grafica' visti sopra, del tipo *tunnel*), franc. *milieu* → ant. it. *miluogo* 'centro' (*lieu* = luogo), ingl. *pocketbook* → sved. *pocketbok* 'libro tascabile' (*book* = bok 'libro').
- 4) **semicalco**: uno degli elementi del modello è tradotto, l'altro è invece del tutto ignorato e sostituito con materiale morfologico o lessicale proprio della lingua d'arrivo. Soprattutto le lingue slave sono ricche di semicalchi: lo sloveno *gostilna* 'trattoria' dipende dal ted. *Gasthaus*, che però è ricalcato solo nella prima parte (*gost* = Gast 'ospite'), mentre il morfo *-ilna* non ha nulla a che fare con la 'casa' (ted. Haus). Altri esempi tedeschi → sloveni: *Eisenbahn* → *železnica* 'ferrovia' (Eisen = železo 'ferro'), *Mittelpunkt* → *središče* 'centro' (Mittel = sreda 'mezzo'). È interessante il fatto che il ted. *Zeitschrift* 'giornale, rivista' sia stato reso in sloveno sia con un calco perfetto: *časopis* (*čas* = Zeit 'tempo', *pis* = Schrift 'scritto') sia con un semicalco: *časnik* (-nik è un suffisso nominale molto produttivo).

Non mancano casi di **calchi per falsa motivazione**, cioè calchi basati su un'analisi sbagliata degli elementi che compongono il modello straniero, sul loro fraintendimento. Un tipico esempio è quello fornito dal ceco *králík* 'coniglio', che è propriamente un diminutivo di *král* 're'; esso ricalca il medio alto ted. *küniklîn*, analizzato dai parlanti cechi come formato da *künik* (ted. mod. *König*) 're' + il suff. dim. *-lîn* (ted. mod. *-lein*). In realtà la parola tedesca non ha niente a che fare con un 'piccolo re', ma è un prestito dal latino *cuniculus*, da cui deriva anche l'it. *coniglio*.

Calco semantico

Abbiamo questo tipo di calco allorché, date due parole appartenenti a lingue diverse ma legate tra loro da somiglianza formale e/o da un significato-base comune, una delle due acquisisce una nuova accezione che in precedenza era propria solo dell'altra. Anche qui il concetto sarà meglio chiarito dagli esempi: l'it. *realizzare* (un prestito settecentesco dal franc. *réaliser*) fino al Novecento ha avuto solo il significato di 'attuare, effettuare'; da qualche decennio a questa parte però esso viene usato anche nel senso di 'comprendere, rendersi conto'. Si tratta di un calco semantico sull'ingl. *to realize* (pur esso francesismo), che tra le sue accezioni ha per l'appunto quella di 'to understand'. Grazie al prestigio e alla forza di penetrazione di cui da tempo gode la lingua inglese, *to realize* ha agito da modello di riferimento per *realizzare*, 'proiettando' sul suo omologo italiano un significato che gli era proprio. Ugualmente l'it. *sofisticato* significava un tempo solo 'adulterato' (*vino sofisticato*); l'accezione di 'ricercato' (*stile sofisticato, donna sofisticata*) è dovuta a un calco semantico sull'inglese *sophisticated*. Finora abbiamo citato casi di parole in cui entra in gioco la somiglianza formale, ma questa non è necessaria; perché scatti il meccanismo del calco semantico può essere sufficiente la condivisione di un significato di base. L'it. *farfalla* nel senso di 'cravattino' è ricalcato sul franc. *papillon*. Il ted. *Ente* e il franc. *canard* vogliono dire 'anatra' e figuratamente 'notizia giornalistica falsa'; questa seconda accezione del tedesco è ovviamente sorta per un calco semantico sul francese.

LA CLASSIFICAZIONE DELLE LINGUE

Uno dei compiti più importanti della linguistica è quello di **classificare**, ossia di ordinare sistematicamente secondo criteri scientifici, le migliaia e migliaia di lingue esistenti al mondo, esattamente come la botanica fa con le piante o la zoologia con gli animali. La classificazione delle lingue si basa essenzialmente su due diversi parametri, 1) l'**affinità** (o **parentela**) **genetica** e 2) l'**affinità tipologica**. Se scegliamo il primo parametro avremo una **classificazione genealogica** (p. es. l'italiano è una lingua neolatina), se scegliamo il secondo avremo una **classificazione tipologica**, che può essere **morfologica** (p. es. l'italiano è una lingua flessiva analitica) oppure **sintattica** (p. es. l'italiano è una lingua SVO).

A) Parentela genetica

Due (o più) lingue si dicono geneticamente apparentate quando sia scientificamente provato che, in una fase precedente della loro evoluzione, esse facevano parte di uno stesso sistema linguistico. Talvolta la parentela genetica è attestata dalla documentazione scritta giunta fino a noi: è il caso ben noto delle lingue neolatine (o romanze) derivate dal latino (anche se, a dire il vero, non si tratta del latino classico, quello di Cicerone e Virgilio, bensì del latino parlato del Basso Impero, molto meno bene documentato) o delle lingue moderne dell'India derivate dal sanscrito. Altre volte tale parentela ci è assicurata dalle fonti storiche: p. es. sappiamo che l'attuale inglese discende dalla lingua parlata da gruppi di Angli, Sassoni e Juti provenienti dalla Germania settentrionale e sbarcati in Britannia nel V sec. d. C., ed è pertanto imparentato con le altre lingue germaniche. Ma in molti casi, soprattutto quando si tratta di lingue extraeuropee, non sappiamo assolutamente nulla della loro storia e degli eventuali rapporti che le legano (o legavano in epoca antica) con altre lingue; in tali situazioni, per stabilire la parentela genetica è giocoforza ricorrere esclusivamente all'analisi linguistica. Quali sono i requisiti necessari che debbono avere due lingue perché si possa considerarle geneticamente apparentate? Cominciamo subito col dire che, per quanto sia ovvio che lingue parenti condividano gran parte del loro fondo lessicale, il criterio basato sulla quantità di lessemi in comune è quello più infido e malsicuro. Infatti conseguentemente ai molteplici contatti intercorrenti tra realtà linguistiche diverse, ai fenomeni di adstrato e superstrato (v. pag. 6), non è affatto raro che una lingua assuma prestiti massicci di elementi lessicali allogloti. Chiunque studia l'inglese si sarà accorto che moltissime parole (se non la maggioranza) sono di origine latina, un'eredità lasciata dai Normanni che conquistarono l'isola nell'XI sec. e che parlavano francese; ciò non significa che l'inglese sia in qualche modo classificabile come lingua neolatina. Parimenti il lessico turco dei secoli passati era composto per tre quarti da parole di origine araba o persiana, grazie all'assoluto predominio culturale di queste lingue nell'impero ottomano; nondimeno il turco non è affatto imparentato né con l'arabo né col persiano. Ordunque perché un'ipotesi di parentela genetica abbia validità scientifica essa deve soddisfare altri e più sicuri requisiti, tra i quali i principali sono:

1) **Lingue geneticamente apparentate devono presentare corrispondenze fonetiche ricorrenti e prevedibili.**

Vediamo subito un esempio ponendo a confronto latino e lingue romanze:

lat. *nocte(m)* > it. *notte*, fr. *nuit*, spagn. *noche*, port. *noite*, rum. *noapte*;
lat. *lacte* > it. *latte*, fr. *lait*, spagn. *leche*, port. *leite*, rum. *lapte*;
lat. *directu(m)* > it. *diritto*, fr. *droit*, spagn. *derecho*, port. *direito*, rum. *drept*.

Dai termini succitati si evince chiaramente che al nesso latino *ct* corrispondono in modo ricorrente l'italiano *tt*, il francese e portoghese *it*, lo spagnolo *ch*, il rumeno *pt*. Il fatto che tali corrispondenze siano ricorrenti comporta evidentemente anche la loro prevedibilità; infatti dato il lat. *factum* potremmo ipotizzare con sufficiente certezza che ne derivino it. *fatto*, fr. *fait*, spagn. *hecho*, port. *feito*, rum. *fapt*, anche se putacaso queste parole non fossero effettivamente attestate. Quando un mutamento fonetico si presenta con regolarità e prevedibilità esso costituisce una **legge fonetica**. L'affermazione «il gruppo consonantico *ct* del latino diventa in italiano *tt*, in francese e portoghese *it*, in spagnolo *ch*, in rumeno *pt*» è una legge fonetica.

L'esistenza delle leggi fonetiche permette una delle operazioni capitali della linguistica storico-comparativa, vale a dire la **ricostruzione**. Una volta cioè assodato a livello sincronico che un gruppo di lingue è legato da parentela genetica, possiamo risalire diacronicamente a fasi anteriori del loro sviluppo ricostruendo (nei limiti del possibile) la lingua (o protolingua) da cui derivano. P. es. la legge di Grimm¹ afferma che in base alla seconda rotazione consonantica una dentale occlusiva sorda germanica diventa un'affricata (o fricativa, a seconda dei casi) in alto tedesco. Confrontando quindi l'ingl. *ten* (< ags. *tīen*) e il ted. *zehn* (< ant. alto ted. *zehan*) è l'inglese (così come il gotico, l'ant. nordico, l'ant. sassone) a mantenere l'occlusiva dentale originaria; procedendo con questo sistema per gli altri suoni, si può ricostruire per il numerale 'dieci' una forma germanica comune **tehun*. Se poi compariamo il germ. **tehun* con il sanscr. *daśa-*, il gr. *déka*, il lat. *decem* e le tante altre lingue che formano la famiglia indoeuropea, è ricostruibile un 'indoeuropeo comune' (ma sono etichette da usare sempre con cautela²) **dek'm*, che dà ragione - foneticamente - delle forme delle varie lingue, in quanto la *d* indoeuropea rimane tale in sanscrito, greco, latino e diventa *t* in germanico, la *m* sonante (*m*) diventa *a* in sanscrito e greco, *em* in latino, *un* in germanico (cfr. il numerale per 'sette': ie. **septm* > sanscr. *sapta-*, gr. *heptá*, lat. *septem*, germ. **sebun*), ecc.

2) Lingue geneticamente apparentate devono presentare una serie di morfemi della stessa origine.

Se prendiamo la frase inglese *the professor clearly explained Grimm's importance for the history of German philology* 'il professore ha chiaramente spiegato l'importanza di Grimm per la storia della filologia tedesca' noteremo che i lessemi che la compongono sono quasi tutti d'origine latina (precisamente da *professor*, *clarus*, *explanare*, *importantia*, *historia*, *germanus*, *philologia* [grecismo mediato dal latino]). Che cosa impedisce allora di considerare l'inglese una lingua neolatina? Innanzitutto i termini che non appartengono al lessico culturale, ma a quello di base (cioè l'articolo *the* e le preposizioni *for*, *of*) sono germanici e non latini, ma soprattutto i tre morfemi grammaticali presenti nella frase sono anch'essi d'origine germanica: il suffisso avverbiale *-ly* è la forma abbreviata dell'antico *-lice* che proviene a sua volta dal suffisso aggettivale *-lic* (cfr. ted. *-lich*); il 'simple past tense' in *-ed* corrisponde perfettamente al preterito debole tedesco (*-te*), olandese (*-de*), danese (*-(e)de*), svedese (*-(a)de*), ecc.; il suffisso genitivale in *-s* (il cosiddetto 'genitivo sassone') si ritrova tale e quale in tutte le lingue germaniche.

Ponendo a confronto i sistemi grammaticali delle lingue indoeuropee salta subito agli occhi una quasi perfetta identità tra i loro morfemi sia nominali sia verbali. Per fare una semplice esemplificazione, diamo qui di seguito il quadro delle desinenze verbali primarie delle prime tre persone quale risulta da alcune fra le lingue più importanti e di più antica attestazione:

¹ Vedi G. Graffi, *Breve storia della linguistica*, pp. 117-118.

² E' in corso da parecchi anni una vivace polemica tra diversi linguisti riguardo alla questione dell'indoeuropeo ricostruito: è davvero possibile una simile ricostruzione, e se sì, essa riflette una lingua realmente parlata o è una pura astrazione? Tralascio poi un altro gravissimo problema: se anche è ricostruibile il *significante* di un segno linguistico, molto spesso non lo è (o lo è in misura assai minore) il suo *significato*. Basterà un solo esempio: il sanscr. *rāj-*, il lat. *rēx*, il celt. **rīg-* 're' ci consentono di ricostruire un ie. **rēg'jo-*, ma su che cosa volesse dire questa parola si possono avanzare solo ipotesi; il fatto che le voci derivate da questo **rēg'jo-* significhino 're' non ci autorizza a proiettare lo stesso significato a una realtà sociale e culturale diversa, risalente a millenni prima.

	sanscrito	greco	(ant.) latino	gotico	ant. slavo
1. pers. sing.	mi	mi	m	m	mǐ
2. pers. sing.	si	si	s	s	si
3. pers. sing.	ti	ti	t	t/th	tǐ

Agli inizi del '900 una spedizione archeologica in Asia Minore scoprì, durante alcuni scavi presso il villaggio di Boğazköy, un enorme numero di tavolette di argilla, che rappresentavano l'archivio ufficiale dell'impero degli Ittiti, fiorito nel corso del 2° millennio a.C. Le tavolette erano scritte in caratteri cuneiformi, ben noti e leggibili - il cuneiforme era il sistema di scrittura usato da tutti i popoli dell'area mesopotamica, dai Sumeri agli Assiro-Babilonesi - ma la lingua in cui erano vergate risultava sconosciuta. Il merito della sua decifrazione spetta al linguista ceco Bedřich Hrozný, il quale, pur essendo un semitista, per primo intuì il carattere indoeuropeo di quella lingua e ne diede notizia nel 1915. Quando poi emerse che le prime tre persone del sistema verbale ittita terminavano rispettivamente in *-mi*, *-si*, *-tsi* (insieme a moltissime altre corrispondenze), fu definitivamente provato che l'ittita apparteneva alla famiglia indoeuropea.

Le principali famiglie linguistiche

Le lingue legate tra loro da parentela genetica sono raggruppabili in **famiglie linguistiche**, secondo il principio della **classificazione genealogica**. Si tratta, com'è ovvio, di una classificazione sorta in analogia con quella naturalistica: le 'famiglie' di lingue corrispondono - per logica e metodo - alle 'famiglie' di animali o piante. Ciascuna famiglia può poi suddividersi, come vedremo, in gruppi, sottogruppi e rami. Diamo ora un elenco delle principali famiglie linguistiche, generalmente riconosciute e accettate dalla comunità degli studiosi.

N.B. Ai fini dell'esame, bisogna sapere soltanto le famiglie linguistiche e lingue stampate in grassetto.

1. **Indoeuropeo**

La famiglia indoeuropea (denominazione che risale all'inizio dell'Ottocento) comprende un grandissimo numero di lingue, quasi tutte le europee e molte di quelle asiatiche, ed è comunemente suddivisa in 10 gruppi (alcuni formati da una sola lingua):

a) **Indoiranico**

Si suddivide in due sottogruppi, indiano e iranico. La lingua indiana più antica è il vedico, in cui a partire dal XIII sec. a.C. vennero redatti i *Veda*, i testi sacri dell'induismo; da esso si sviluppò il **sanscrito**, ossia la lingua «perfezionata, elaborata» (tale è il significato di *saṃskṛta-*) dai grammatici (in particolare Pāṇini) nella quale sono composti i grandi poemi epici (*Rāmāyaṇa* e *Mahābhārata*) e tutta la letteratura classica indiana. Dai dialetti parlati del sanscrito (i cosiddetti *pracriti*) sorsero le lingue medioindiane (tra le quali il pali, la lingua dei testi buddhisti) e infine le lingue indiane moderne, di cui citiamo almeno l'**hindi** (la più diffusa, scritta nell'alfabeto sanscrito, il *devanāgarī*; una sua variante è l'urdu, parlato in Pakistan e scritto coi caratteri arabi) e il bengali, lingua di R. Tagore, premio Nobel per la letteratura nel 1913.

Il sottogruppo iranico si divide a sua volta in due rami, uno occidentale e uno orientale. Al primo appartengono l'antico persiano (documentato dalle iscrizioni cuneiformi dei re Achemenidi, VI-V sec. a.C.) e l'avestico, cioè la lingua dell'*Avesta*, il libro sacro del mazdeismo o zoroastrismo, la religione di Zarathustra. In epoca successiva troviamo il medio persiano (o *pahlavi*) fino a giungere alle lingue neoiraniche, la più importante delle quali è il **persiano moderno**. Quanto alle lingue iraniche orientali, ricordiamo il pashtu, parlato in Afghanistan.

b) Armeno

Le prime attestazioni dell'armeno, lingua parlata nel Caucaso meridionale e nella Turchia orientale, risalgono al V sec. d.C. e sono costituite quasi esclusivamente da testi cristiani. Questo «armeno classico» è rimasto in uso fino ai giorni nostri come lingua liturgica della chiesa cristiana armena (uno dei centri principali si trova a Venezia, nell'isola di San Lazzaro).

L'armeno moderno (considerevolmente diverso da quello antico) si suddivide in due varietà, l'armeno occidentale e l'armeno orientale. Su quest'ultimo si basa la lingua ufficiale dell'odierna Repubblica d'Armenia.

c) Greco

È superfluo rammentare quale sia stata l'importanza storica e culturale della lingua greca nel mondo antico. I primi Greci giunsero nel bacino dell'Egeo probabilmente intorno all'inizio del secondo millennio a.C.; le prime documentazioni della lingua da loro parlata ci sono fornite da un discreto numero di tavolette d'argilla rinvenute a Creta, incise in una scrittura sillabografica chiamata lineare B e databili tra il 1450 e il 1200 a.C. La decifrazione della lineare B, avvenuta soltanto poco dopo il 1950 per merito dell'inglese Michael Ventris, permise di stabilire senza alcun dubbio che la lingua delle tavolette, definita miceneo, era una forma di greco molto antica. Le attestazioni micenee precedono di parecchi secoli quelle del greco storico, che appare frazionato in vari dialetti, tra i quali vanno ricordati lo ionico-attico (quello più importante per la letteratura), il dorico e l'eolico. A partire dal IV sec. a.C. dallo ionico-attico si sviluppò la *koiné* («(dialetto) comune») che si diffuse in tutto il mondo ellenizzato ad opera di Alessandro Magno e dei suoi successori. Dalla *koiné* deriva il greco medievale o bizantino che a sua volta si evolve nel greco moderno o neogreco.

d) Albanese

Le prime fonti che ci documentano sull'albanese risalgono appena al XV sec. d.C. e la prima opera stampata in tale lingua, una traduzione del Messale Romano, appare solo nel 1555. Attualmente, oltre che in Albania, esso è parlato in Kosovo, in Macedonia, in qualche zona greca e in diverse colonie dell'Italia meridionale. Si divide in due varietà, il ghego a nord e il toscano a sud. L'odierno albanese standard è basato sul toscano.

Nel corso della storia l'albanese ha subito un forte influsso, soprattutto lessicale, latino (e più tardi anche slavo, turco, neogreco e italiano), ma conserva un fondo molto antico. Le sue origini sono ancora discusse: alcuni studiosi ritengono che continui l'antico illirico, altri il tracio, lingue del primo millennio a.C., entrambe ben poco conosciute.

e) Italico

Con questo termine si indica comunemente il complesso delle lingue indoeuropee parlate nell'Italia antica (quindi l'etrusco, la cui parentela con altre lingue è ancora discussa, non fa parte dell'italico). Purtroppo gran parte di queste lingue sono attestate in modo estremamente frammentario, sicché è assai arduo stabilire i loro rapporti reciproci. In genere le lingue italiche sono suddivise in due gruppi, quello osco-umbro e quello **latino**, anche se molti studiosi pensano che questi gruppi siano giunti in Italia separatamente, in epoche e per vie diverse, e che perciò un originario «italico comune» non sia mai esistito.

Ad ogni modo, come si sa, ben presto il latino soprafecce tutte le altre lingue, e, grazie all'espansione dell'Impero romano, si diffuse ben al di là del suolo italiano. Il latino parlato nelle varie regioni dell'Impero (chiamato anche, meno bene, latino volgare) diede origine alle odierne lingue **neolatine** o **romanze**: l'**italiano**, il **provenzale** od **occitano** (la lingua d'oc degli antichi trovatori), il **francese**, lo **spagnolo** o **castigliano**, il **catalano**, il **portoghese**, il **romeno**. Inoltre ci sono anche lingue neolatine minori come il sardo, il romancio (una delle lingue ufficiali della Svizzera) e il dalmatico (estinto; l'ultimo parlante morì a Veglia, isola del Quarnero, nel 1898).

f) **Celtico**

Le popolazioni celtiche occupavano anticamente gran parte dell'Europa occidentale, arrivando ad insediarsi sin nei Balcani e nell'Asia Minore. La conquista romana e le migrazioni dei Germani ridussero poi drasticamente il loro territorio. Il gruppo delle lingue celtiche si suddivide in celtico continentale o gallico (la lingua degli antichi Galli, poi sommersa dal latino) e **celtico insulare**, diviso a sua volta in **britannico** e **gaelico**. Al ramo britannico appartengono il **cimrico** o **gallese** (lingua del Galles, accanto all'inglese), il cornico (parlato in Cornovaglia fino al XVIII sec., ora estinto) e il bretone, parlato nella Bretagna francese, dove venne importato dai Britanni in fuga dall'Inghilterra nel VI sec. d.C. (quindi non si tratta di un continuatore del celtico continentale). Al gaelico appartengono l'**irlandese**, lingua ufficiale della Repubblica d'Irlanda, lo scozzese (in via d'estinzione, chiamato anche gaelico scozzese per distinguerlo dallo scozzese inteso come varietà d'inglese parlato in Scozia) e il mancese, parlato sull'isola di Man fino a pochi decenni fa e ora scomparso.

g) **Germanico**

Le lingue germaniche sono tradizionalmente ripartite in tre grandi sottogruppi: il **germanico orientale**, il **germanico occidentale** e il **germanico settentrionale**. Del germanico orientale ci è rimasta testimonianza di una sola lingua, il **gotico**, documentato soprattutto dalla traduzione di parte della Bibbia fatta dal vescovo Wulfila nel IV sec. d.C., che per la sua antichità ha una grandissima importanza linguistica.

Nel germanico occidentale si possono distinguere due raggruppamenti, l'anglo-frisone e il **tedesco** in senso lato. L'anglo-frisone è formato dal frisone, parlato nella parte settentrionale dei Paesi Bassi, la Frisia, e nelle isole adiacenti, e dall'**inglese**, che si può dividere cronologicamente in tre periodi: 1) **antico inglese** o **anglosassone** (dal VII all'XI sec.) con un lessico ancora omogeneo e prevalentemente germanico; 2) **medio inglese** (dal XII al XV sec.; la sua massima espressione letteraria è costituita dall'opera di Geoffrey Chaucer), quando, con la conquista normanna, si ha una massiccia penetrazione di vocaboli francesi e una notevole semplificazione della morfologia; 3) **inglese moderno**, caratterizzato da una fortissima alterazione fonetica (nel vocalismo abbiamo il cosiddetto *great vowel shift*), mentre la grafia resta sostanzialmente quella del medio inglese.

Il tedesco si divide territorialmente in **basso tedesco** (Nord) e **alto tedesco** (Sud). Al basso tedesco appartengono l'antico sassone, lingua in cui è scritto lo *Heliand* (un poema di quasi 6000 versi, IX sec.) e il basso francone, dal quale si sono sviluppati, in epoca moderna, l'**olandese** e il **fiammingo**, che si servono di una lingua scritta comune chiamata **neerlandese**. Ricordiamo anche l'afrikaans, sorto dalla lingua parlata dai coloni olandesi (i boeri) in Sud Africa (v. pag. 21). L'**alto tedesco** si contrappone linguisticamente a tutte le altre lingue germaniche per la presenza della seconda rotazione consonantica o *zweite Lautverschiebung* (v. sopra p. 12, n. 1). Anch'esso si divide in tre periodi: 1) **antico alto tedesco** (*althochdeutsch*; 750-1050 ca.), frammentato in numerosi dialetti; 2) **medio alto tedesco** (*mittelhochdeutsch*; 1050-1350 ca.), la lingua della grande civiltà cortese e cavalleresca, del *Nibelungenlied* e del *Minnesang*; 3) **tedesco moderno** (*neuhochdeutsch*), lingua ufficiale di Germania, Austria e Svizzera.

Il germanico settentrionale in epoca antica è attestato da una lingua sostanzialmente unitaria definita **antico nordico**, nella quale sono redatte numerose e importanti opere letterarie come p. es. i famosi poemi dell'*Edda*. Solo dopo il XIII sec. cominciarono a svilupparsi autonomamente le singole lingue nordiche, vale a dire l'**islandese**, di gran lunga quella più conservativa - dal punto di vista morfologico l'islandese odierno non differisce praticamente dall'antico nordico -, il **norvegese**, il **danese**, lo **svedese**, il faroese o feringio (lingua ufficiale delle Fær Øer accanto al danese).

h) **Baltoslavo**

Questo gruppo, formato dai sottogruppi **baltico** e **slavo**, non è ammesso da tutti gli studiosi; secondo alcuni le somiglianze tra baltico e slavo sarebbero dovute semplicemente a un secolare contatto fra i parlanti di queste lingue.

Il baltico comprende il **lituano**, il **lettone** e l'antico prussiano (oggi estinto). Le lingue baltiche, nonostante siano attestate in epoca molto tarda (non prima del XVI sec.) conservano un aspetto piuttosto arcaico e sono molto importanti negli studi di indoeuropeistica.

I documenti slavi più antichi risalgono al IX sec. d. C. e sono costituiti da traduzioni bibliche e altre opere di contenuto religioso. La lingua in cui sono redatti viene chiamata **antico slavo ecclesiastico**. Le lingue slave moderne si dividono in 1) **slavo meridionale**, con **bulgaro**, **macedone**, **serbo**, **croato** e **sloveno**; 2) **slavo occidentale**, con **ceco**, **slovacco**, **polacco** e sorabo; 3) **slavo orientale**, con **russo**, **bielorosso** e **ucraino**.

i) **Anatolico**

Per questo gruppo (formato da lingue tutte estintesi già in era precristiana) ci limitiamo a citare l'**ittita** (v. sopra, pag. 13), che è anche la lingua indoeuropea di più antica attestazione (le prime iscrizioni risalgono agli inizi del secondo millennio a. C.) ed è quindi molto importante ai fini degli studi indoeuropeistici.

j) **Tocario**

Verso la fine dell'Ottocento vennero alla luce in monasteri buddisti del Turkestan orientale (appartenente alla Cina) molti manoscritti antichi redatti in varie lingue sia conosciute sia sconosciute; ma anche queste ultime erano leggibili perché scritte in caratteri noti. Una delle lingue sconosciute, i cui testi risalivano al VI-VIII sec. d.C., si rivelò di chiara matrice indoeuropea e ad essa fu dato il nome di tocario, poiché si pensò, probabilmente in modo errato, che essa fosse la lingua dei Tocari, antica popolazione di stirpe iranica citata da molteplici fonti (greche, sanscrite, cinesi, turche). Il tocario si suddivide in due varietà dette tocario A (orientale) e tocario B (occidentale).

2. **Afroasiatico (camitosemitico)**

La famiglia **afroasiatica** (un tempo definita anche **camitosemitica**) comprende i seguenti gruppi:

- a) **egiziano**, una delle lingue del mondo di più antica documentazione - le prime iscrizioni geroglifiche risalgono al IV millennio a.C. - il cui stadio di sviluppo più tardo è costituito dal **copto**, che fu la lingua degli Egiziani cristianizzati e si spense con la conquista araba, rimanendo in uso solo come lingua liturgica della chiesa copta;
- b) berbero, comprendente le lingue parlate dai tuareg e da altre popolazioni sahariane;
- c) cuscitico, al quale appartengono numerose lingue dell'Etiopia e della Somalia, tra le quali la più importante è il somalo;
- d) **semitico**, suddiviso in vari sottogruppi: 1) semitico orientale, rappresentato dall'accadico, antichissima lingua degli Assiro-Babilonesi, scritta in caratteri cuneiformi; 2) semitico nordoccidentale, attestato in epoca antica soprattutto da fenicio, **aramaico** (la lingua di Gesù) ed **ebraico** (in cui è scritto l'Antico Testamento). Sull'ebraico biblico è modellato l'**ebraico moderno** o **ivrit**, la lingua ufficiale dello stato d'Israele; 3) semitico sudoccidentale, comprendente un ramo centrale, con l'**arabo**, e un ramo meridionale, con diverse lingue etiopiche, tra cui l'amarico.

Alcuni studiosi ascrivono a questa famiglia anche altre lingue africane, quelle che formano i gruppi nilotico e ciadico.

3. Uralico

Di questa famiglia fanno parte il piccolo gruppo delle lingue samoiede, nella Siberia orientale, e il vasto raggruppamento delle lingue **ugrofinniche**, la cui parentela fu riconosciuta fin dal '700. Tra le lingue ugrofinniche annoveriamo l'**ungherese**, il lappone, il **finlandese** o **finnico** e l'**estone** (le ultime due appartengono al sottogruppo **baltofinnico**); inoltre, divise in vari sottogruppi, parecchie altre lingue minori parlate in diverse zone della Russia sia europea sia asiatica.

4. Altaico

E' formato da tre gruppi:

- a) **mongolo**, comprendente varie lingue parlate in Mongolia, Cina, Siberia e Manciuria;
- b) **manciu-tunguso**, costituito dal mancese o manciù, in Manciuria, e dalle lingue tunguse, ormai ridotte a pochi parlanti, nella Siberia orientale;
- c) **turco**, vastissimo gruppo diffuso in larga parte dell'Asia e le cui prime attestazioni risalgono all'VIII sec. d.C. Tra le principali lingue turche moderne ricordiamo il **turco** propriamente detto, la lingua odierna della repubblica di Turchia (fino al XIX secolo definita più propriamente **turco-ottomano** od **osmanli**), il turkmeno, il kirghiso, l'usbeco, l'uiguro.

Alcuni studiosi ritengono che le lingue uraliche e quelle altaiche siano apparentate e postulano quindi una famiglia uralo-altaica, negata invece da altri. Va ancora poi detto che qualche linguista non crede neppure alla parentela fra mongolo, manciu-tunguso e turco, mettendo con ciò in discussione l'esistenza di una famiglia altaica.

5. Lingue caucasiche

Con questa etichetta si indicano le numerose lingue parlate nel Caucaso, ad eccezione di quelle appartenenti alla famiglia indoeuropea (come l'armeno) o a quella altaica (p.es. l'azero, parlato in Azerbaigian, è una lingua turca). Si tratta quindi di una designazione di valore più geografico che linguistico, in quanto non si è potuta ancora stabilire con certezza una loro parentela originaria. E' vero che le lingue caucasiche mostrano un discreto numero di corrispondenze sul piano sia fonetico sia morfologico (p. es. sono quasi tutte lingue ergative), ma tali tratti comuni potrebbero dipendere dal contatto e dall'influsso reciproco, cioè da fenomeni di adstrato. La più importante (e anche l'unica di attestazione antica) tra le lingue caucasiche è il **georgiano**, lingua ufficiale della Repubblica di Georgia.

6. Lingue dell'Asia e del Pacifico

Oltre a quelle già citate, nell'area dell'Asia e dell'Oceania troviamo principalmente quattro grandi famiglie, la **sino-tibetana**, la **dravidica**, l'**austro-asiatica** e la **maleo-polinesiana**.

- a) Le lingue più importanti della famiglia **sino-tibetana**, che viene suddivisa in vari gruppi e sottogruppi, sono: il **cinese** (distinto in numerose varietà, tra loro assai differenziate; l'odierno cinese letterario standard, basato sul dialetto di Pechino, è chiamato **cinese mandarino**), il tibetano, il birmano, il thai o thailandese, il lao o laotiano.
- b) E' probabile che le lingue **dravidiche** fossero diffuse in tutta l'India all'epoca dell'invasione degli indoeuropei nel II millennio a.C. Attualmente esse ricoprono compattamente la sola parte meridionale dell'India e le principali sono il tamil, il kannada o canarese, il telugu, tutte con ricca letteratura fin da epoca medievale.
- c) La famiglia **austro-asiatica** è divisa in due gruppi, le lingue munda (India orientale) e le lingue mon-khmer, tra le quali ricordiamo il khmer o cambogiano e il vietnamita (ma alcuni ritengono che quest'ultimo sia un idioma «misto», sorto dalla fusione di una lingua thai con una mon-khmer).

- d) La famiglia **maleo-polinesiana** ha un'estensione immensa (si va dalle Hawaii alla Nuova Zelanda, dal Madagascar all'isola di Pasqua) e si divide in tre gruppi (ognuno con un elevatissimo numero di lingue, di cui citiamo solo qualcuna): 1) lingue **indonesiane**, che comprendono il **malese-indonesiano** (cioè le lingue ufficiali della Malaysia e dell'Indonesia, che con la riforma ortografica del 1972 si sono in pratica unificate, tranne qualche differenza lessicale), il tagalog (Filippine), il giavanese (Giava), il malgascio (Madagascar); 2) lingue **polinesiane**, parlate negli arcipelaghi della Polinesia (Hawaii, Samoa, Figi, Tonga, ecc.), alle quale appartiene anche il maori (Nuova Zelanda); 3) lingue **melanesiane**, nelle isole della Melanesia (le Salomone, le Bismarck, le Nuove Ebridi, ecc.).

Le lingue maleo-polinesiane sono caratterizzate da un sistema fonologico molto semplice (le vocali sono quelle fondamentali, le consonanti poche e dei tipi più comuni) con parole formate per lo più da sillabe aperte, secondo la struttura CVCV. Nonostante l'enorme territorio su cui sono disseminate, esse presentano inoltre sorprendenti corrispondenze nel lessico di base. A titolo di esempio, citiamo i numerali 'due', 'tre' e 'cinque' in malgascio e samoano (si noti che il Madagascar e le Samoa sono separate da molte migliaia di chilometri):

	'due'	'tre'	'cinque'
malgascio	<i>rua</i>	<i>telu</i>	<i>limi</i>
samoano	<i>lua</i>	<i>tolu</i>	<i>lima</i>

Per completare il quadro, va detto che le circa duecento lingue dell'Australia (gran parte delle quali in via d'estinzione) sono ancora poco indagate e una loro presunta parentela originaria resta *sub judice*.

Infine segnaliamo che per due importanti lingue asiatiche, il giapponese e il coreano, non si è stabilito alcun chiaro rapporto con altre famiglie linguistiche, nonostante svariati tentativi in tal senso (alcuni studiosi li considerano appartenenti alla famiglia altaica).

7. Lingue dell'Africa

Abbiamo già visto che molte lingue dell'Africa centrosettentrionale appartengono alla famiglia afroasiatica, le rimanenti si ripartiscono in parecchie famiglie con numerosi gruppi e sottogruppi. Per non appesantire troppo il discorso, citeremo qui soltanto l'importante famiglia **bantu**, che occupa quasi tutta l'Africa meridionale e la cui lingua principale è il **swahili**, lingua ufficiale di Tanzania, Uganda e Kenia, ma nota e usata come lingua veicolare anche negli stati vicini. Le lingue bantu sono caratterizzate da un complesso sistema di classi nominali, segnalate da appositi prefissi (v. il testo del Simone, p. 175).

8. Lingue delle Americhe

Usando gli stessi criteri adoperati per l'indoeuropeo, si è calcolato che vi siano 54 famiglie linguistiche nell'America settentrionale, 23 nel Messico e nell'America centrale e circa 75 (addirittura 108 secondo altri) nell'America meridionale. Essendo ovviamente impossibile elencare in questa sede tali famiglie, accenneremo invece alla proposta di E. Sapir, che nel 1929 classificò le lingue del Nordamerica in sei grandi famiglie, sulla base di lontane somiglianze strutturali: eschimoaleutino (di cui fanno parte i dialetti parlati dagli Eschimesi o Inuit), na-dene (Alaska, Canada nordoccidentale), algonchino (Montagne Rocciose, Grandi Laghi), hoka-sioux (California, Dakota, fino al Texas e al Nicaragua), penuti (Oregon, California), uto-azteco (a cui appartiene il nahuatl, la lingua degli Aztechi).

Più recentemente J. Greenberg, per classificare genealogicamente le lingue, ha introdotto il procedimento della «comparazione di massa» (*mass comparison*), che si basa sull'identificazione di elementi lessicali e morfologici: se un certo elemento di una lingua è paragonabile a quello di un'altra,

lo si ritiene un tratto ereditario. Con questo criterio Greenberg ha drasticamente ridotto a tre il numero di famiglie americane: l'eschimo-aleutino, il na-dene (già individuati da Sapir, come s'è detto) e l'amerindiano, che comprenderebbe tutte le lingue restanti. La classificazione di Greenberg ha però dato adito a grandi controversie ed è respinta da molti studiosi.

Per quanto riguarda l'America meridionale, ci limiteremo a ricordare il nome della sua lingua indigena più importante e diffusa, il **quechua** (grafia spagnola; all'italiana checiua), la lingua degli Incas, oggi parlata in Perù, Ecuador, Bolivia e nelle aree limitrofe di Argentina, Cile e Colombia.

B) Affinità tipologica

Due o più lingue si dicono tipologicamente affini quando presentano **caratteristiche funzionali e strutturali simili**, indipendentemente dalla dimensione spaziale e temporale. Non ha alcuna importanza che gli elementi fonologici, morfologici o sintattici comparati abbiano un'origine comune, quello che conta è la funzione che essi svolgono all'interno del rispettivo sistema linguistico. L'affinità tipologica perciò accomuna lingue anche lontanissime tra loro in termini storici e geografici e senza alcun contatto reciproco. P. es. l'ungherese e il somalo appartengono a famiglie diverse (l'uralica e l'afroasiatica, come abbiamo visto) ed ovviamente i loro parlanti non hanno mai avuto contatti. In entrambe però è presente il fenomeno dell'armonia vocalica e questo tratto fonologico comune le rende tipologicamente affini.

La **linguistica tipologica** si occupa quindi, in base all'affinità sopra descritta, di individuare dei **tipi linguistici**, ossia modelli di sistema linguistico, e di classificare le lingue secondo tali tipi. Mentre la classificazione genealogica opera su base diacronica, attraverso la ricostruzione e l'indagine sulla preistoria delle lingue, la classificazione tipologica ha un'impostazione sincronica, basata su metodi di comparazione. Benché ovviamente gli elementi messi a confronto possano riguardare anche la fonologia (v. sopra l'esempio dell'armonia vocalica), le classificazioni tipologiche più studiate e diffuse (e solo di queste ci occuperemo) sono quella **morfologica** e quella **sintattica**.

1. Tipologia morfologica

Dato che i tipi morfologici sono trattati da R. Simone nei *Nuovi Fondamenti di Linguistica* (pp. 95-97, che invitiamo gli studenti a leggere attentamente), qui ci limiteremo a qualche breve nota integrativa.

Per osservare in pratica come funziona una lingua **isolante** analizziamo la frase cinese *wǒ gěi tā mǎi le shū* {io + DAT. + lui/lei + comprare + PASS. + libro} 'gli/le ho comprato un libro': *wǒ* è il pronome di 1° persona singolare, *gěi* può fungere da morfema lessicale col valore di 'dare' oppure (come in questo caso) da morfema grammaticale di dativo, equivalente alle nostre preposizioni 'a' o 'per', *tā* è il pronome di 3° persona singolare, *mǎi* 'comprare', *le* morfema che indica azione passata, *shū* 'libro'. Come si può notare, manca totalmente qualsiasi traccia di flessione, ogni parola è immodificabile; le marche grammaticali sono date da morfi liberi o preposti (*gěi*, dativo) o posposti (*le*, passato). Inoltre l'ordine sintattico è necessariamente rigido, la funzione di soggetto di *wǒ* e quella di oggetto di *shū* - così come la funzione di tutti gli altri elementi - si desumono unicamente dalla loro posizione all'interno della frase; qualsiasi spostamento sintattico altererebbe il significato della frase o la renderebbe incomprensibile. Mancando i processi di derivazione, l'unica risorsa a disposizione del cinese per formare parole nuove è la composizione: p. es. unendo le parole *zhōng* 'centro, in mezzo' e *guó* 'paese' otteniamo *Zhōngguó* 'Cina' (che per i Cinesi è il 'paese di mezzo'); aggiungendo ancora *rén* 'persona' avremo *zhōngguórén* 'Cinese'. Molte parole del linguaggio tecnico-scientifico cinese sono formate col lessema *diàn* 'elettricità', p. es. *diànhuà* {elettricità + parlare} 'telefono', *diànnǎo* {elettricità + cervello} 'computer', *diàntī* {elettricità + scala} 'ascensore'.

Per esemplificare la differenza fra lingue **flessive** e **agglutinanti** mettiamo a confronto la declinazione delle parole latine *lingua* e *opus* con quella dei loro omologhi turchi *dil* ‘lingua’ e *eser* ‘opera’:

Singolare				
Nominativo	<i>lingua</i>	<i>opus</i>	<i>dil</i>	<i>eser</i>
Genitivo	<i>linguae</i>	<i>operis</i>	<i>dilin</i>	<i>eserin</i>
Dativo	<i>linguae</i>	<i>operī</i>	<i>dile</i>	<i>esere</i>
Accusativo	<i>linguam</i>	<i>opus</i>	<i>dili</i>	<i>eseri</i>
Ablativo	<i>linguā</i>	<i>opere</i>	<i>dilden</i>	<i>eserden</i>
Plurale				
Nominativo	<i>linguae</i>	<i>opera</i>	<i>diller</i>	<i>eserler</i>
Genitivo	<i>linguārum</i>	<i>operum</i>	<i>dillerin</i>	<i>eserlerin</i>
Dativo	<i>linguīs</i>	<i>operibus</i>	<i>dillere</i>	<i>eserlere</i>
Accusativo	<i>linguās</i>	<i>opera</i>	<i>dilleri</i>	<i>eserleri</i>
Ablativo	<i>linguīs</i>	<i>operibus</i>	<i>dillerden</i>	<i>eserlerden</i>

Salta subito agli occhi che in una lingua flessiva come il latino le declinazioni sono fortemente differenziate: non c'è alcuna corrispondenza tra le desinenze casuali di *lingua* e quelle di *opus*, né tanto meno, all'interno della stessa declinazione, tra quelle di singolare e plurale (fra *ae* e *arum* non v'è alcuna somiglianza, benché entrambe marchino il genitivo). Al contrario il turco, agglutinante, mostra un unico modello di flessione, estremamente regolare e valido sia per il singolare sia per il plurale (marcato dal morfo *ler*; ovviamente ciascun morfo ha vari allomorfi dovuti all'armonia vocalica). Le desinenze latine si fondono spesso con la vocale tematica distruggendo i confini di morfo, mentre quelle turche sono sempre facilmente separabili dalla radice. Infine in latino, come in tutte le lingue flessive, vari morfemi grammaticali si accumulano nello stesso morfo (l'*am* di *linguam* esprime inscindibilmente l'accusativo, il singolare e il femminile, ossia il caso, il numero e il genere), laddove nelle lingue agglutinanti tra morfema e morfo c'è un rapporto biunivoco, cioè ogni morfema ha un morfo proprio (vedi il turco, dove non esiste un 'genitivo plurale' analogo al latino, ma una marca *ler* che indica il plurale, e solo quello, e una marca *in* che indica il genitivo, e solo quello).

Le lingue **incorporanti** (ma oggi si preferisce definirle **polisintetiche**) sono caratterizzate da un'agglutinazione spinta all'estremo dei vari morfemi grammaticali e lessicali dell'enunciato, fino ad arrivare alla cosiddetta 'parola-frase'. Esempi particolarmente significativi di polisintesi sono quelli offerti dalle lingue nordamericane, come il seguente, che è tratto dal paiute (una lingua parlata nello Utah, appartenente alla famiglia uto-azteca): *wiitokuchumpukurüganinyugwicantüim* (naturalmente non tutte le 'parole' del paiute hanno questa eccezionale lunghezza), che significa 'quelli che andranno a sedersi e a tagliare a pezzi con un coltello una mucca nera'.

Un'importante distinzione concernente la tipologia morfologica è quella tra lingue **sintetiche** (o **fusive**), che tendono a combinare più morfemi in un'unica parola, e lingue **analitiche**, dove molte funzioni grammaticali vengono espresse da morfemi liberi e ognuna delle unità resta relativamente indipendente dalle altre. L'indice più alto di sintesi è rappresentato dalle lingue polisintetiche, quello più basso dalle lingue isolanti. L'opposizione sintetico/analitico si può riscontrare soprattutto all'interno dei tipi agglutinante e flessivo. P. es. il turco e il giapponese sono entrambi agglutinanti; vediamo però come la frase *l'insegnante ha dato il libro al bambino* si traduce in queste lingue:

turco	<i>öğretmen</i>	<i>çocuğa</i>	<i>kitabı</i>	<i>verdi</i>
	insegnante-NOM	bambino-DAT	libro-ACC	ha dato
giapponese	<i>sensei ga</i>	<i>kodomo ni</i>	<i>hon o</i>	<i>kudasaimashita</i>

Come risulta evidente, mentre il turco appare pienamente sintetico, il giapponese mostra chiari tratti analitici, evidenziati dall'uso di morfemi liberi posposti (*ga* per il nominativo, *ni* per il dativo, *o* per l'accusativo) per marcare il caso.

Nel corso della loro evoluzione, le lingue flessive tendono in generale a trasformarsi da sintetiche in analitiche: il latino e l'anglosassone erano sintetici, le lingue neolatine e l'inglese sono analitici. La perdita dei tratti sintetici può spingersi tanto oltre da avvicinare certe lingue al tipo isolante, che, come s'è detto, è quello massimamente analitico. L'esempio migliore in questo senso è fornito dall'afrikaans, lingua che fa parte del gruppo germanico (v. pag. 15), dove s'è verificata una scomparsa quasi totale dei morfemi legati insieme alla tendenza a ridurre le parole a monosillabi (cfr. *ek kon my oë nie glo nie* 'non potevo credere ai miei occhi'). Il sistema verbale (tranne pochi residui di preteriti forti) è ridotto a due uniche forme: prendendo come modello il verbo per 'leggere', abbiamo *lees*, usato per l'infinito, l'imperativo e tutte le persone del presente, e *gelees*, participio che preceduto dall'ausiliare *het* 'avere' esprime i tempi del passato; non solo, ma i verbi composti con certi prefissi (*be-*, *ont-* e altri) non prendono neppure questo *ge-*, sicché hanno un'unica forma valida per tutti i tempi e modi. Pertanto, dal punto di vista tipologico, verbi afrikaans come *besit* 'possedere', *ontdek* 'scoprire' si comportano esattamente come verbi cinesi.

In conclusione va sottolineato come i confini che separano un tipo morfologico dall'altro non sono assolutamente netti e definiti. Il turco, ad esempio, che viene additato come lingua agglutinante *par excellence*, può presentare parole 'monstre' quali *şöhretlendirmediklerinizdendir* 'è uno di quelli che non avete reso famosi', che si attribuirebbero facilmente a una lingua polisintetica. Allo stesso modo lingue flessive presentano spesso tratti agglutinanti e viceversa. Insomma ben difficilmente le lingue rappresentano tipi morfologici **puri**, bensì tipi in maggiore o minore misura **misti**; se una lingua si definisce flessiva, ciò significa che i suoi tratti sono prevalentemente flessivi rispetto a quelli che potrebbero accomunarla a un altro tipo.

2. Tipologia sintattica e universali

Il fondatore della tipologia sintattica moderna è il già citato Joseph H. Greenberg (1915-2001), linguista statunitense che ne pose le basi nel 1963 con la pubblicazione dell'articolo intitolato 'Alcuni universali della grammatica con particolare riferimento all'ordine degli elementi significativi' (*Some Universal of Grammar with particular reference to the order of meaningful elements*).

Vengono definiti **universali linguistici** le caratteristiche condivise da tutte (o quasi) le lingue naturali, indipendentemente dalle loro reciproche relazioni storiche; p. es. affermazioni quali 'non esiste sistema fonologico che non comprenda sia vocali sia consonanti' oppure 'categorie sintattiche come il nome e il verbo sono presenti nella struttura di ogni lingua nota' sono degli universali. Nel lavoro citato, Greenberg si è proposto di dedurre una serie di universali di natura grammaticale in base ad un'indagine basata essenzialmente su osservazioni sintattiche. A tale scopo, uno dei suoi criteri è stato quello di esaminare **l'ordine relativo di soggetto, verbo e oggetto in frasi dichiarative con soggetto e oggetto nominali**. Per spiegarci meglio, facciamo un esempio traducendo la frase *il gatto* (soggetto, S) *ha mangiato* (verbo, V) *il topo* (oggetto, O) in inglese, turco e arabo:

inglese	<i>the cat (S) ate (V) the mouse (O)</i>
turco	<i>kedi (S) fareyi (O) yedi (V)</i>
arabo	<i>akala (V) l-qitṭu (S) l-fa'ra (O)</i>

Pertanto, mentre l'inglese ha un ordine SVO (come quello italiano), il turco ce l'ha SOV ('il gatto il topo ha mangiato') e l'arabo VSO ('ha mangiato il gatto il topo'). Comparando le varie lingue del mondo, Greenberg ha notato che quasi tutte hanno uno di questi tre ordini; gli altri teoricamente possibili (OSV, OVS, VOS) sono rarissimi o addirittura inesistenti e sono caratterizzati dal fatto che in essi l'oggetto precede il soggetto. Ciò ha permesso a Greenberg di formulare il suo primo universale (ordine dominante significa ordine normale, non marcato):

Nelle frasi dichiarative con oggetto e soggetto nominali, l'ordine dominante è quasi sempre quello in cui il soggetto precede l'oggetto.

Oltre a quello dell'ordine basico, Greenberg prende in considerazione altri due criteri: 1) **l'uso da parte di una lingua di preposizioni oppure posposizioni**. P. es. in latino *cum* è normalmente una preposizione (*cum amicis meis* 'con i miei amici'), ma con certi pronomi personali viene posposto (*tecum* 'con te', *nobiscum* 'con noi') diventando così una posposizione. In genere una lingua usa o preposizioni o posposizioni; 2) **la posizione degli aggettivi qualificativi in rapporto al nome**. P. es. nelle lingue germaniche l'aggettivo precede il nome, in quelle neolatine lo segue (sempre che abbia valore qualificativo, come in *uomo buono*; nel caso di *buon uomo* l'aggettivo ha valore attributivo).

Scegliendo una campionatura di trenta lingue³ e combinando i tre criteri Greenberg ha ottenuto la seguente tabella:

	SVO	SOV	VSO
Pr-A+N	4	0	0
Pr-N+A	6	0	6
Po-A+N	1	6	0
Po-N+A	2	5	0

(Pr = uso di preposizioni, Po = uso di posposizioni, A + N = l'aggettivo precede il nome, N + A = il nome precede l'aggettivo).

Dalla tabella risulta quindi che tra le lingue SVO dieci sono preposizionali (quattro con l'ordine aggettivo + nome, sei il contrario) e tre posposizionali (una con l'ordine aggettivo + nome, due il contrario). Più interessanti sono i dati riguardanti gli altri tipi: tutte le lingue SOV appaiono posposizionali, tutte le VSO preposizionali (e con ordine N + A). Va tuttavia detto che, al di fuori delle trenta lingue campione, ce n'è qualcuna SOV preposizionale (p. es. persiano e amarico) e qualcuna VSO con ordine A + N. Sulla base di tali risultanze Greenberg ha potuto formulare gli universali seguenti:

Le lingue con l'ordine dominante VSO sono sempre preposizionali.

Con una frequenza di gran lunga più che casuale, le lingue con l'ordine normale SOV sono posposizionali.

Con frequenza di gran lunga più che casuale, le lingue con l'ordine dominante VSO hanno l'aggettivo dopo il nome.

³ Sette europee (italiano, norvegese, finlandese, gallese, serbo, greco moderno, basco), sette africane (yoruba, nubico, swahili, fulani, masai, songhai, berbero), nove asiatiche (turco, ebraico, burushaski, hindi, kannada, giapponese, thailandese, birmano, malese), cinque americane (maya, zapoteco, quechua, chibcha, guarani), due dell'Oceania (maori, loritja). Il basco e il burushaski (parlato nel Caracorum, regione del Kashmir) sono lingue isolate, non apparentate con alcuna famiglia nota. Di queste trenta lingue tredici sono SVO (italiano, norvegese, finlandese, greco moderno, serbo, yoruba, fulani, swahili, songhai, thailandese, malese, maya, guarani), undici SOV (basco, nubico, turco, burushaski, hindi, kannada, giapponese, birmano, quechua, chibcha, loritja), sei VSO (gallese, ebraico, masai, berbero, zapoteco, maori).

Con l'aggiunta di numerosi altri criteri (posizione del genitivo, delle particelle interrogative, dell'ausiliare, ecc.), Greenberg arriva ad enunciare 45 universali, di cui 25 sono puramente sintattici, mentre gli altri prendono in considerazione anche aspetti morfologici. Molti di questi universali sono **implicazionali**, assumono cioè la forma «dato x in una particolare lingua, si trova sempre y », p. es.:

Se una lingua ha l'ordine dominante SOV e il genitivo segue il nome reggente, allora allo stesso modo l'aggettivo segue il nome.

Se l'oggetto pronominale segue il verbo, l'oggetto nominale fa altrettanto.

Se una lingua presenta flessione, presenta sempre derivazione.

Se una lingua ha la categoria del genere, ha sempre la categoria del numero.

L'approccio tipologico di Greenberg è stato molto criticato, soprattutto da parte dei generativisti, i linguisti della scuola chomskiana. Tuttavia, a nostro giudizio, tali critiche sono ingenerose; allo studioso americano va riconosciuto il merito di aver indagato e confrontato, con acume e acribia, un numero incredibile di lingue, ottenendone risultati di rilievo. La sua tipologia dell'ordine basico è ormai diventata un 'classico' e le etichette SVO, SOV, VSO vengono usate comunemente da tutti. Certo, la tipologia sintattica è ancora relativamente 'giovane' rispetto a quella morfologica, che ha buoni due secoli di storia alle spalle; ma è prevedibile che essa verrà sempre più utilizzata, anche nel campo della linguistica diacronica (p. es. non mancano recenti studi sull'ordine basico [SOV?] dell'indoeuropeo ricostruito).