

Classificazione delle lingue

Uno dei compiti più importanti della linguistica è quello di **classificare**, ossia di ordinare in modo sistematico secondo criteri scientifici, le migliaia di lingue del mondo, così come la botanica fa con le piante o la zoologia con gli animali. La classificazione delle lingue si basa essenzialmente su due parametri:

A. l'**affinità** (o **parentela**) **genetica**. In base a questo parametro si ha una **classificazione genealogica** (p. es. l'italiano è una lingua *romanza* o *neolatina*).

B. l'**affinità tipologica**. In base a questo secondo parametro si ha 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 imparentate quando sia scientificamente provato che, in una fase precedente della loro evoluzione, esse facevano parte di uno stesso sistema linguistico.

La parentela genetica si può basare su più fatti:

- La documentazione scritta giunta fino a noi: è il caso delle lingue romanze (o neolatine) risalenti al latino (anche se non si tratta del latino classico, quello di Cicerone e Virgilio, bensì del latino parlato del Basso Impero, molto meno documentato), o delle lingue moderne dell'India derivate dal sanscrito.

- Le fonti storiche: p. es. l'attuale inglese discende dalla lingua parlata dagli Anglo-Sassoni (che consistevano in 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, si sa poco o nulla della loro storia e degli eventuali rapporti che le legano o hanno legate con altre lingue; in tali situazioni, per stabilire la parentela genetica bisogna ricorrere esclusivamente all'analisi linguistica.

Quali sono i requisiti perché due lingue possano essere considerate geneticamente imparentate?

È vero che due o più lingue imparentate condividono gran parte del loro fondo lessicale; tuttavia, il criterio che si basa sulla quantità di lessemi in comune tra due o più lingue è quello più infido e meno sicuro. Infatti, come conseguenza di molteplici contatti avvenuti tra realtà linguistiche completamente diverse, non è affatto raro che una lingua assuma un numero cospicuo di prestiti consistenti in elementi lessicali provenienti da altre lingue. Per esempio, moltissime parole della lingua inglese sono di origine latina: si tratta di un'eredità lasciata dai Normanni, i quali conquistarono l'isola nell'XI sec. e parlavano francese; ciò non significa, ovviamente, che l'inglese possa essere classificabile come lingua romanza. Allo stesso modo, 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; ma il turco non è imparentato né con l'arabo né con il persiano.

Pertanto, affinché un'ipotesi di parentela genetica sia scientificamente valida, deve soddisfare requisiti più sicuri, tra i quali i principali sono:

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

Come esempio si può fare un confronto tra latino e alcune lingue romanze (italiano, francese, spagnolo, portoghese e romeno):

lat. *nocte(m)* > it. *notte*, fr. *nuit*, sp. *noche*, port. *noite*, rom. *noapte*;

lat. *lacte* > it. *latte*, fr. *lait*, sp. *leche*, port. *leite*, rom. *lapte*;

lat. *directu(m)* > it. *diritto*, fr. *droit*, sp. *derecho*, port. *direito*, rom. *drept*.

Da questi termini si deduce che al nesso latino *ct* corrispondono in modo ricorrente l'italiano *tt*, il francese e portoghese *it*, lo spagnolo *ch* e il romeno *pt*. Il fatto che tali corrispondenze siano ricorrenti comporta anche che esse sono prevedibili: è possibile ipotizzare con sufficiente certezza che il lat. *factum* dia it. *fatto*, fr. *fait*, port. *feito*, sp. *hecho*, rom. *fapt*, anche qualora queste parole non fossero effettivamente attestate. Quando un mutamento fonetico si presenta con regolarità e prevedibilità, esso costituisce una **legge fonetica**, per cui in questo caso si può affermare: "il gruppo consonantico *ct* del latino diventa in italiano *tt*, in francese e portoghese *it*, in spagnolo *ch* e in romeno *pt*".

L'esistenza delle leggi fonetiche permette una delle operazioni fondamentali della linguistica storico-comparativa, ovvero la **ricostruzione**. Una volta cioè assodato a livello sincronico che un gruppo di lingue è legato da parentela genetica, è possibile risalire diacronicamente a fasi anteriori del loro sviluppo ricostruendo (nei limiti del possibile) la lingua (o protolingua) da cui queste lingue derivano.

P. es. in base alla seconda rotazione consonantica, già descritta da Grimm, una dentale occlusiva sorda germanica diventa un'affricata (o fricativa, a seconda dei casi) in alto tedesco. Confrontando quindi l'ingl. *ten* (< anglosass. *tīen*) e il ted. *zehn* (< ant. alto ted. *zehan*) è l'inglese (così come il gotico, l'antico nordico, l'antico sassone) a mantenere l'occlusiva *t* originaria; procedendo con questo sistema per gli altri suoni, si può ricostruire per il numerale "dieci" una forma germanica comune **tehun*. Inoltre, comparando il germ. **tehun* con il sanscr. *daśa-*, il gr. *δέκα (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 cautela¹) **dékṃ-*, che dal punto di vista fonetico spiega le forme delle varie lingue, in quanto la *d* indoeuropea rimane tale in sanscrito, greco, latino e diventa *t* in germanico; la *m* sonante (*ṃ*) diventa *a* in sanscrito e greco, *em* in latino, *un* in germanico (cfr. il numerale per "sette": indoeur. **septṃ-* > sanscr. *sapta-*, gr. *ἑπτὰ (heptá)*, lat. *septem*, germ. **sebun*), ecc.

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

Si può prendere come esempio la seguente 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". I lessemi che la compongono sono quasi tutti di origine latina (< lat. *professor*, *clarus*, *explanare*, *importantia*, *historia*, *germanus*, *philologia*, quest'ultimo un grecismo mediato dal latino). Perché quindi l'inglese non si può considerare una lingua romanza? 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 di origine germanica: il suffisso avverbale *-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*); infine, il suffisso genitivale in *-s* (il cosiddetto "genitivo sassone") appartiene a tutte le lingue germaniche.

Confrontando tra di loro i sistemi grammaticali delle lingue indoeuropee si nota subito una quasi perfetta identità tra i loro morfemi sia nominali sia verbali. Come esempio si possono prendere le desinenze verbali primarie delle prime tre persone di alcune tra le lingue più importanti e di più antica attestazione:

¹ Da diversi anni è in corso 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? A ciò si aggiunge un ulteriore problema: se il *significante* di un segno linguistico può essere ricostruibile, molto spesso il suo *significato* non lo è (o lo è in misura assai minore). Basterà un solo esempio: il sanscr. *rāj-*, il lat. *rēx* e il celt. **rīg-* "re" consentono di ricostruire un indoeur. **rēgjo-*, ma su che cosa volesse dire questa parola si possono avanzare solo ipotesi; il fatto che le voci derivate da questo **rēgjo-* significhino "re" non vuol dire che sia ovvio 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	sǐ
3. pers. sing.	ti	ti	t	t/th	tǐ

Agli inizi del Novecento una spedizione archeologica in Asia Minore scoprì, durante alcuni scavi presso il villaggio di Boğazköy, un centinaio di km a est di Ankara, un enorme numero di tavolette di argilla, la quali rappresentavano l'archivio ufficiale dell'impero degli Ittiti, fiorito nel corso del 2° millennio a. C. Le tavolette erano scritte in caratteri cuneiformi (il cuneiforme è noto in quanto era il sistema di scrittura usato da tutti i popoli dell'area mesopotamica, dai Sumeri agli Assiro-Babilonesi) ma la lingua in cui erano scritte 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.

Il modo principale per ordinare i vari sistemi linguistici del mondo è raggrupparli in **famiglie linguistiche**.