

RAPPORTI TRA LINGUE E MATEMATICA: ESPERANTO E OLTRE

Parte 1. Biiettività fonematica

CESCO REALE

Associazione Mondiale di Esperanto, rappresentante ONU

Festival di Giochi Matematici (www.tuttoenumero.it)

1 Introduzione

Questo articolo è la trascrizione e l'ampliamento dell'omonima conferenza tenuta nel 2011 nell'ambito di *Pianeta Galileo*, progetto promosso dal Consiglio Regionale della Toscana. Lo scopo è quello di mostrare come si possano individuare vari strumenti matematici applicabili allo studio delle lingue che regolano la loro struttura.

Nella conferenza sono stati analizzati i seguenti elementi: biiettività fonematica, biiettività morfologica, associatività e commutatività semantica, opposto e complemento logico, intersezione di insiemi fonetici e semantici nella scrittura cinese, le lingue logiche, linguaggio e metalinguaggio, paradosso 'del barbiere', frasi bisenso, scomposizione semantica in fattori primi, la *Characteristica Universalis* di Leibniz, l'Esperanto come sistema formale (morfemi-assiomi e parole-teoremi), sistemi numerici.

In questo articolo si analizzerà solo il primo tema, espandendolo con analisi che sarebbero troppo lunghe per una conferenza.

2 Biiettività fonematica

Nella parola "gialli" si pronuncia più debolmente la prima o la seconda "i"? A questa domanda il pubblico risponde in genere in maniera eterogenea. Dopo aver fatto ascoltare varie volte una registrazione della parola, c'è chi propende per la prima lettera, chi per la seconda, chi per la parità. L'audio è accompagnato dalla visione di un grafico: il segnale temporale che rappresenta l'onda acustica (Figura 1).

A questo punto si fa ascoltare una registrazione della parola temporalmente invertita, anch'essa accompagnata dal grafico corrispondente (Figura 1). La maggior parte delle persone nota facilmente che in questa seconda registrazione si sente una sola "i", quella all'inizio della parola. Aggiungiamo la trascrizione fonetica: [i l l a ʒ d], ecco i foni che ascoltiamo.

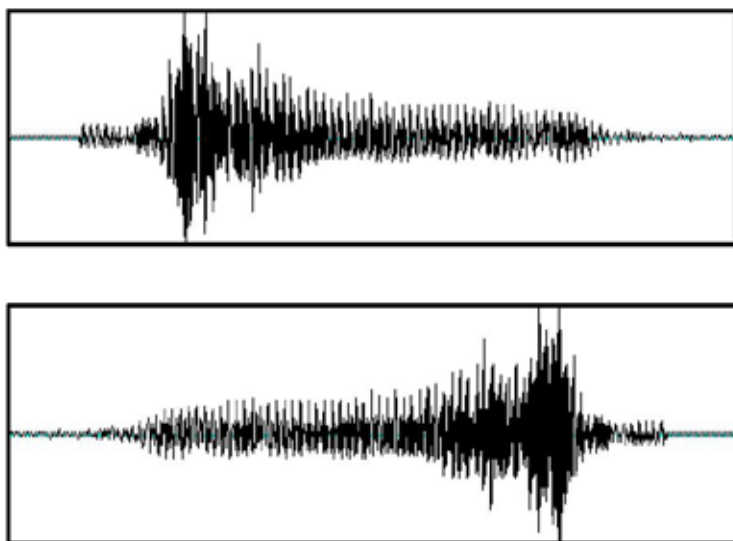


Figura 1.

(Ad essere precisi, il suono della [d] finale è quasi inudibile, per renderlo udibile faccio ascoltare una seconda registrazione in cui ho fatto precedere alla parola “gialli” una breve vocale [ə], registrazione che poi ho invertito temporalmente.) Se però invertiamo temporalmente i foni, essi restano gli stessi pur cambiando l’ordine: quindi in realtà solo una “i” viene pronunciata! Questa è la seconda “i” della parola “gialli”. La prima invece è solo un simbolo grafico per indicare che la “g” non è da pronunciare come occlusiva velare /g/ (come in “galli”), ma come affricata postalveolare /dʒ/.

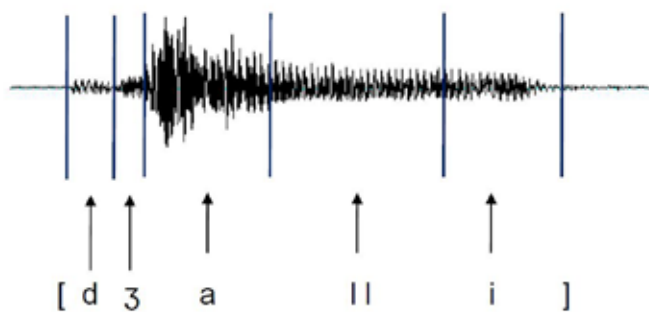


Figura 2.

È, in effetti, credenza comune tra gli italofofoni che la lingua italiana si legga così come si scrive. Indubbiamente, la trasparenza fonematica dell’Italiano è superiore a quella di molte altre lingue quali Inglese, Francese, Olandese. Ciò non vuol dire che questa corrispondenza tra lettere e fonemi sia perfettamente biunivoca (biunivoca).

Vediamone un esempio:



Figura 3.

2.1 Relazione lettere-fonemi

Proviamo allora ad analizzare la relazione esistente in italiano tra lettere e fonemi: definiamo un dominio L (lettere) e un codominio F (fonemi). La relazione diretta $G: L \rightarrow F$ corrisponde alla lettura, mentre la relazione inversa G^{-1} corrisponde alla scrittura. Ci rendiamo conto che non c'è iniettività né in un senso (lettere $\rightarrow$ fonemi) né nell'altro (fonemi $\rightarrow$ lettere). G quindi non è una funzione in quanto esistono elementi del dominio che hanno più di un'immagine (ad esempio, la "c"). Essa è dunque una multifunzione. Inoltre è non iniettiva in quanto esistono elementi del codominio i quali hanno più di una controimmagine (ad esempio, /k/). Notiamo inoltre che, per definire tale relazione G su tutto il dominio, dobbiamo includere nel codominio il fonema nullo / /, immagine della H, oltre che in alcuni casi della I, come abbiamo appena visto.

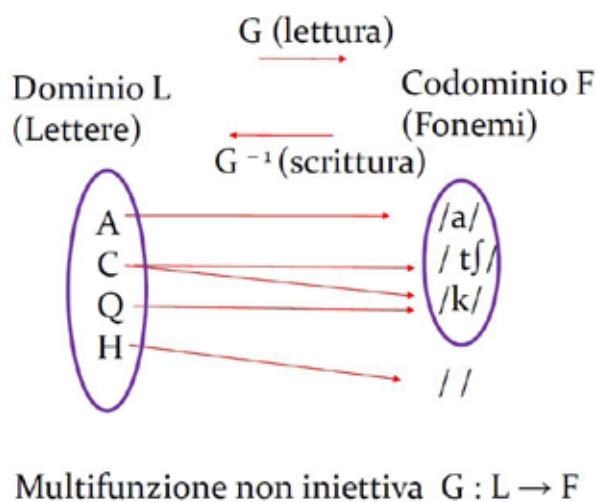


Figura 4.

Il fatto che in entrambi i sensi non ci sia iniettività – e quindi che G non sia una funzione – è una caratteristica scomoda non solo in matematica, ma anche in linguistica. Infatti, restando per ora nell'ambito della fonetica, è molto più facile imparare una lingua se la sua pronuncia è regolare, in particolare se la sua lettura è univocamente definita (funzione) e ancor più se è biettiva.

L'Italiano, nonostante la sua fonetica sia regolata da una multifunzione non iniettiva, è comunque (complessivamente) abbastanza regolare. Altre lingue sono sicuramente meno regolari. Facciamo due esempi tra le lingue a noi più note: Francese e Inglese.

- Francese. Fermo restando che ogni lettera (tranne Z) può essere pronunciata con un valore fonetico che ha anche in italiano, una parola come CHOIX è ben lontana da /koiks/ ma si pronuncia /ʃwa/. Questa è anzi una parola regolare, in quanto in francese molto spesso CH $\rightarrow$ /ʃ/, OI $\rightarrow$ /wa/, X finale $\rightarrow$ /. Ma esistono poi moltissime eccezioni alle regole, per cui possiamo trovare CH $\rightarrow$ /k/, X finale $\rightarrow$ /s/, o ancora A o L mute (SAOUL $\rightarrow$ /su/), F muta (CEUFS $\rightarrow$ /ø/), ecc.;
- Inglese. Esso è ancora più irregolare. Facciamo solo qualche esempio di parole che hanno pronunce particolarmente sorprendenti, anche per chi conosce bene la lingua: GAOL $\rightarrow$ /dʒeɪl/, GREENWICH $\rightarrow$ /'ɡrenɪʃ/, CORPS $\rightarrow$ /kɔː/, HICCOUGH $\rightarrow$ /'hɪkəp/, HALFPENNYWORTH $\rightarrow$ /'heɪpəθ/.

Possiamo chiederci se esistano lingue foneticamente regolari, cioè regolate da una funzione biiettiva tra lettere e fonemi. Ebbene sì: ad esempio, il Serbo-croato, il Coreano e il Finlandese sono sistemi fonetici perfettamente o quasi perfettamente biiettivi. L'Esperanto è perfettamente biiettivo: 28 lettere, 28 fonemi, a ogni lettera corrisponde un solo fonema e a ogni fonema una sola lettera.

Nel caso dell'Esperanto, tutta la sua grammatica è stata pianificata per essere facile, logica e regolare. Ma anche nei casi precedenti, se oggi troviamo regolarità è perché in determinati momenti storici è stato fatto un lavoro di pianificazione, regolarizzazione e sistematizzazione per ottenere una trasparenza fonematica.

2.2 Indice di biiettività fonematica

Abbiamo parlato di sistemi fonetici più o meno regolari, trasparenti e biiettivi. È però possibile misurare questo parametro? Proponiamo qui degli indici di regolarità della pronuncia.

Cominciamo dall'indice di biiettività lettere-fonemi per l'Italiano. Definiamo, come sopra, un dominio L delle lettere e un codominio F dei fonemi. L contiene le 21 lettere dell'alfabeto italiano. In generale, è bene lasciare fuori le lettere che compaiono solo nei prestiti, per non falsare l'indice. È chiaro che, se considerassimo anche tutte le pronunce strane che hanno molti prestiti in italiano, la regolarità si abbasserebbe molto, modificando sensibilmente la nostra misura. In F possiamo chiederci se inserire uniti o separati i fonemi in corso di neutralizzazione, come /s/ e /z/. Si è optato per la separazione. Inoltre – come detto – dobbiamo inserire in F il fonema nullo / / per poter definire la relazione su tutto il dominio. Totale: 31 fonemi (7 vocali, 21 consonanti, 2 semiconsonanti, 1 fonema nullo).

Possiamo costruire una matrice M in cui a ogni colonna corrisponde una lettera $\lambda \in L$ e a ogni riga un fonema $\varphi \in F$. A ogni elemento $M(\lambda, \varphi)$ della matrice M attribuiremo il valore 1 se la lettera λ può rappresentare (o concorrere a rappresentare) il fonema φ . Altrimenti gli attribuiremo il valore 0.

In alcuni casi λ rappresenta una sequenza di fonemi (che chiameremo “plurifonema”) contenente φ (esempio in Spagnolo: $X \mapsto /ks/$); in altri casi λ appartiene a una sequenza di lettere (che chiameremo “plurigramma”) e quindi concorre a rappresentare il fonema φ (es. in Italiano: G in $GN \mapsto /n/$) o un plurifonema contenente φ (es. in Francese: $OI \mapsto /wa/$). In tutti questi casi $M(\lambda, \varphi) = 1$. Come detto, la relazione diretta $G: L \longrightarrow F$ è la lettura, la relazione inversa $G^{-1}: F \longrightarrow L$ è la scrittura.

Ora consideriamo il seguente gioco: Ada estrae una parola a caso da un testo con trascrizione fonetica, ne estrae a caso l' i -esima lettera λ e la dice a Ugo, il quale non conosce né il numero i , né la parola, ma ha sotto gli occhi la matrice M . Ugo deve dire un fonema che sia immagine di λ e fa un punto se il fonema proposto è rappresentato in quella parola da λ . Se nella parola c'è un'altra lettera uguale a λ ciò non verrà considerato. Se, ad esempio, la lettera sorteggiata è la prima della parola CACI, Ugo fa un punto solo se dice $/k/$, ma non se dice $/tʃ/$. Definiamo $p(\lambda)$ la probabilità di attribuire correttamente a λ il fonema che λ effettivamente rappresenta in quel caso.

Per il momento consideriamo equiprobabili i possibili fonemi $\varphi = G(\lambda^*)$ associabili al λ^* scelto, cioè tali per cui $M(\lambda^*, \varphi) = 1$. Sotto questa ipotesi, $p(\lambda)$ è il reciproco del numero di tali fonemi. Ad esempio la lettera G in italiano può essere associata a 4 fonemi: $/g/$, $/dʒ/$, $/ɲ/$, $/ʎ/$. Quindi $p(G) = 1/4$.

Definiamo I_{RL} “indice di regolarità in lettura” (ovvero, I_{IS} “indice di iniettività in scrittura”) come la media aritmetica di tutte le $p(\lambda)$ calcolate per le varie lettere nel dominio L .

Possiamo ripetere lo stesso ragionamento scambiando lettere e fonemi e considerando la relazione di scrittura $G^{-1}: F \longrightarrow L$. Dato un fonema φ scelto a caso in una parola scelta a caso, definiamo $p(\varphi)$ la probabilità di attribuire correttamente a φ la lettera da cui φ è effettivamente rappresentato in quel caso. Per il momento consideriamo equiprobabili le possibili lettere $\lambda = G^{-1}(\varphi^*)$ associabili al φ^* scelto, cioè tali per cui $M(\lambda, \varphi^*) = 1$. Sotto questa ipotesi, $p(\varphi)$ è il reciproco del numero di tali lettere. Ad esempio, il fonema $/k/$ in Italiano può essere rappresentato da 2 lettere: C oppure Q . Quindi $p(/k/) = 1/2$. Definiamo I_{RS} “indice di regolarità in scrittura” (ovvero I_{IL} “indice di iniettività in lettura”) come la media aritmetica di tutte le $p(\varphi)$ calcolate per i vari fonemi in F (codominio di G e dominio di G^{-1}). L’“indice di biattività” I_B è definito come la media aritmetica di I_{RL} e I_{RS} .

Riportiamo di seguito i valori calcolati per Francese, Italiano, Spagnolo ed Esperanto e le matrici relative:

	Francese	Italiano	Spagnolo	Esperanto
I_{RL}	44.3%	72.6%	86.7%	100%
I_{RS}	57.2%	91.9%	81.2%	100%
I_B	50.8%	82.3%	83.9%	100%

Figura 5.

Matrice della relazione lettere-fonemi per il francese

Français	A	Æ	B	C	Ç	D	E	F	G	H	I	L	M	N	O	Œ	P	Q	R	S	T	U	V	X	Y	Z				p(φ)		
/a/	1						1				1				1										1			5		0.2		
/ɑ/	1																											1		1		
/ã/	1						1						1	1														4		0.25		
/b/			1																									1		1		
/k/				1													1							1				3		0.333		
/j/				1							1																	2		0.5		
/d/						1																						1		1		
/e/	1	1					1				1					1									1			6		0.167		
/ɛ/	1						1				1														1			4		0.25		
/ē/							1				1		1	1											1			5		0.2		
/ə/	1						1				1																	3		0.333		
/ø/							1									1												2		0.5		
/œ/							1									1												2		0.5		
/f/								1		1							1											3		0.333		
/g/										1														1				2		0.5		
/l/	1					1	1	1			1		1				1			1	1	1		1	1	1		13		0.077		
/v/												1														1		2		0.5		
/j/												1	1												1			3		0.333		
/ʒ/										1																		1		1		
/v/													1															1		1		
/mv/														1														1		1		
/n/															1													1		1		
/ɲ/										1					1													2		0.5		
/o/	1															1							1					3		0.333		
/ɔ/	1															1							1					3		0.333		
/β/													1	1	1													3		0.333		
/p/																	1											1		1		
/ts/																			1									1		1		
/s/				1	1															1	1			1				5		0.2		
/z/																				1				1		1		3		0.333		
/ʀ/																					1					1		1		1		
/w/															1							1						2		0.5		
/v/																							1					1		1		
/hw/											1				1										1			3		0.333		
/y/																							1					1		1		
/u/																							1					1		1		
/œ/													1	1									1					3		0.333		
																															0.443	I _{RS}
	9	1	1	3	1	2	9	2	3	3	8	3	5	6	6	3	3	1	1	3	3	7	1	5	8	2						
p(λ)	0.111		1																													
																															</	

Figura 6.

Matrice della relazione lettere-fonemi per lo spagnolo

[illegible]

Figura 8.

Matrice della relazione lettere-fonemi per l'esperanto

Esperanto	A	B	C	Ĉ	D	E	F	G	Ĝ	H	Ĥ	I	J	Ĵ	K	L	M	N	O	P	R	S	Ŝ	T	U	Ŭ	V	Z		p(φ)			
/a/	1																												1	1			
/b/		1																												1	1		
/ts/			1																											1	1		
/tʃ/				1																										1	1		
/d/					1																									1	1		
/e/						1																								1	1		
/f/							1																							1	1		
/g/								1																						1	1		
/dʒ/									1																					1	1		
/h/										1																				1	1		
/x/											1																			1	1		
/i/												1																		1	1		
/j/													1																	1	1		
/ʒ/														1																1	1		
/k/															1															1	1		
/l/																1														1	1		
/m/																	1													1	1		
/n/																		1												1	1		
/o/																			1											1	1		
/p/																				1										1	1		
/r/																					1									1	1		
/s/																						1								1	1		
/ʃ/																							1							1	1		
/t/																								1						1	1		
/u/																									1					1	1		
/w/																										1				1	1		
/v/																											1			1	1		
/z/																												1		1	1		
		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
p(λ)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1 ^{RS}	
																															1	1	
																															1 ^{RL}	1 ^B	

Figura 9.

2.3 Altri indici di regolarità della pronuncia

A questo punto possiamo estendere lo stesso procedimento ad altri casi. Innanzitutto possiamo variare dominio e/o codominio. Finora abbiamo considerato la relazione lettere-fonemi (indici fonematici), ma potremmo studiare la relazione *lettere-foni* (indici fonetici), ad esempio per misurare quanto sono frequenti le variazioni allofoniche in una lingua. I valori degli indici fonetici non possono essere maggiori di quelli dei rispettivi indici fonematici. L'alfabeto fonetico internazionale è pensato per essere perfettamente biiettivo rispetto alla relazione *grafemi-foni*.

Potremmo inoltre studiare la relazione *plurigrammi-plurifonemi* – come definita nel § 2.2, ove i plurigrammi sono sequenze di lettere e i plurifonemi sono sequenze di fonemi – e considerare i relativi indici, che chiameremo “plurifonematici”. Questo consente di gestire meglio casi come il Francese *OI* |—> /wa/ e di codificare nella matrice le regole di pronuncia che si riferiscono a gruppi di lettere. I valori degli indici plurifonematici tenderanno quindi a essere maggiori di quelli degli indici fonetici, a patto di scegliere opportunamente dominio e codominio. Presentiamo di seguito una matrice plurigrammi-plurifonemi per il Francese: I_{RL} è aumentato da 57.2% a 69.3%, I_{RS} da 44.3% a 80.2%, I_B da 50.8% a 74.8%.

[illegible]

Figura 10.

Possiamo inoltre migliorare l'accuratezza degli indici introducendo delle *informazioni statistiche*. Immaginiamo di analizzare moltissimi testi: ogni volta che troviamo una

lettera λ vediamo a che fonema (o fonemi) φ essa è associata e incrementiamo di 1 l'elemento $M(\lambda, \varphi)$, ottenendo una rappresentazione di frequenza. Detta N la somma finale di tutti i valori della matrice, la $p(\lambda)$ diventa:

$$(1) \quad p(\lambda) = \text{MAX}_{\varphi} \frac{M(\lambda, \varphi)}{N}$$

Infatti, per Ugo la migliore strategia è puntare sempre sul fonema con massima probabilità tra quelli associabili alla lettera scelta.

Detta N_L la cardinalità di L (il numero di lettere) e N_F la cardinalità di F (il numero di fonemi), la probabilità di presentazione della lettera λ è:

$$(2) \quad q(\lambda) = \sum_{\varphi=1}^{N_F} \frac{M(\lambda, \varphi)}{N}$$

e I_{RL} diventa la media delle $p(\lambda)$ pesate rispetto alle rispettive probabilità di presentazione $q(\lambda)$:

$$(3) \quad I_{RL} = \sum_{\lambda=1}^{N_L} p(\lambda) q(\lambda)$$

Analogamente:

$$(4) \quad p(\varphi) = \text{MAX}_{\lambda} \frac{M(\lambda, \varphi)}{N} \quad q(\varphi) = \sum_{\lambda=1}^{N_L} \frac{M(\lambda, \varphi)}{N} \quad I_{RS} = \sum_{\varphi=1}^{N_F} p(\varphi) q(\varphi)$$

I_B resta uguale alla media aritmetica di I_{RL} e I_{RS} . Chiameremo questi nuovi indici “a posteriori”, per distinguerli dai precedenti che chiameremo “a priori”.

3. Conclusioni

Gli indici proposti offrono dei modi per misurare il concetto di *regolarità* della pronuncia nei significati esposti e permettono quindi di comparare lingue diverse mediante questo parametro. Ciò permette di capire quantitativamente perché sistemi fonetici che sono stati parzialmente o totalmente pianificati/sistematizzati sono più logici, regolari e facili da apprendere rispetto ad altri. L'Esperanto ne è un esempio; e il suo indice di biattività fonematica è massimo, cioè 1 (100%).

Nel seguito della conferenza sono stati applicati procedimenti simili per studiare la morfologia e la semantica, mostrando che anche in questi ambiti la pianificazione può migliorare la regolarità e la facilità di apprendimento.

BIBLIO/SITOGRAFIA

- Eco, U., *La ricerca della lingua perfetta*, Laterza, Roma-Bari 1993.
- Migliorini, B., *Manuale di Esperanto*, Paolet, San Vito al Tagliamento 1922.
- Canepari, L., *Introduzione alla fonetica*, Einaudi, Torino 1979.
- Minnaja, C., Lingvistikaj aplikoj de iuj matematikaj teknikoj, in *Miscellanea Interlingvistica*, a cura di I. Szerdahelyi, Tankonykiadó, Budapest 1980, pp. 161-169.
- Dodero, N., Baroncini, P., Manfredi, R., *Lineamenti di matematica. Modulo B. Relazioni. Funzioni. Calcolo letterale. Per i Licei*, Ghisetti e Corvi, Lecce 2000.
- http://it.wikipedia.org/wiki/Lingua_esperanto
- Studi sulla facilità e propedeuticità dell'Esperanto:
<http://easiestlanguage.info/LanguageLearningResearch.html>
- Video introduttivo alla lingua e al mondo esperantista, in 6 capitoli:
<http://www.youtube.com/watch?v=ApF7fWXpQYQ&feature=related>