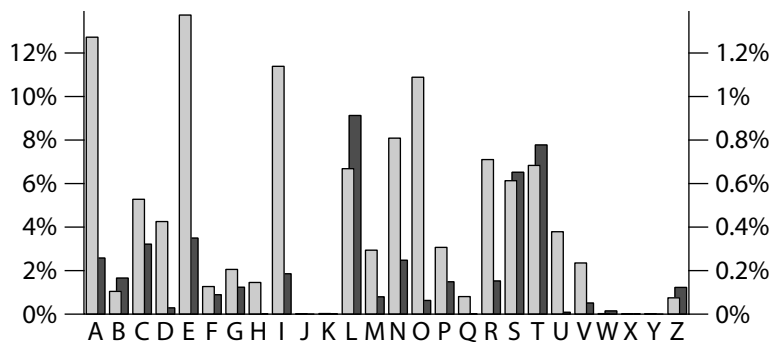


Frequenza dei digrammi in italiano

Utilizzando gli stessi quattro testi di prima (ma questa volta considerati tutt'insieme), studiamo la frequenza delle coppie successive di lettere (digrammi). In quest'analisi consideriamo successive anche lettere separate da punteggiatura, spazi, andate a capo, ecc.

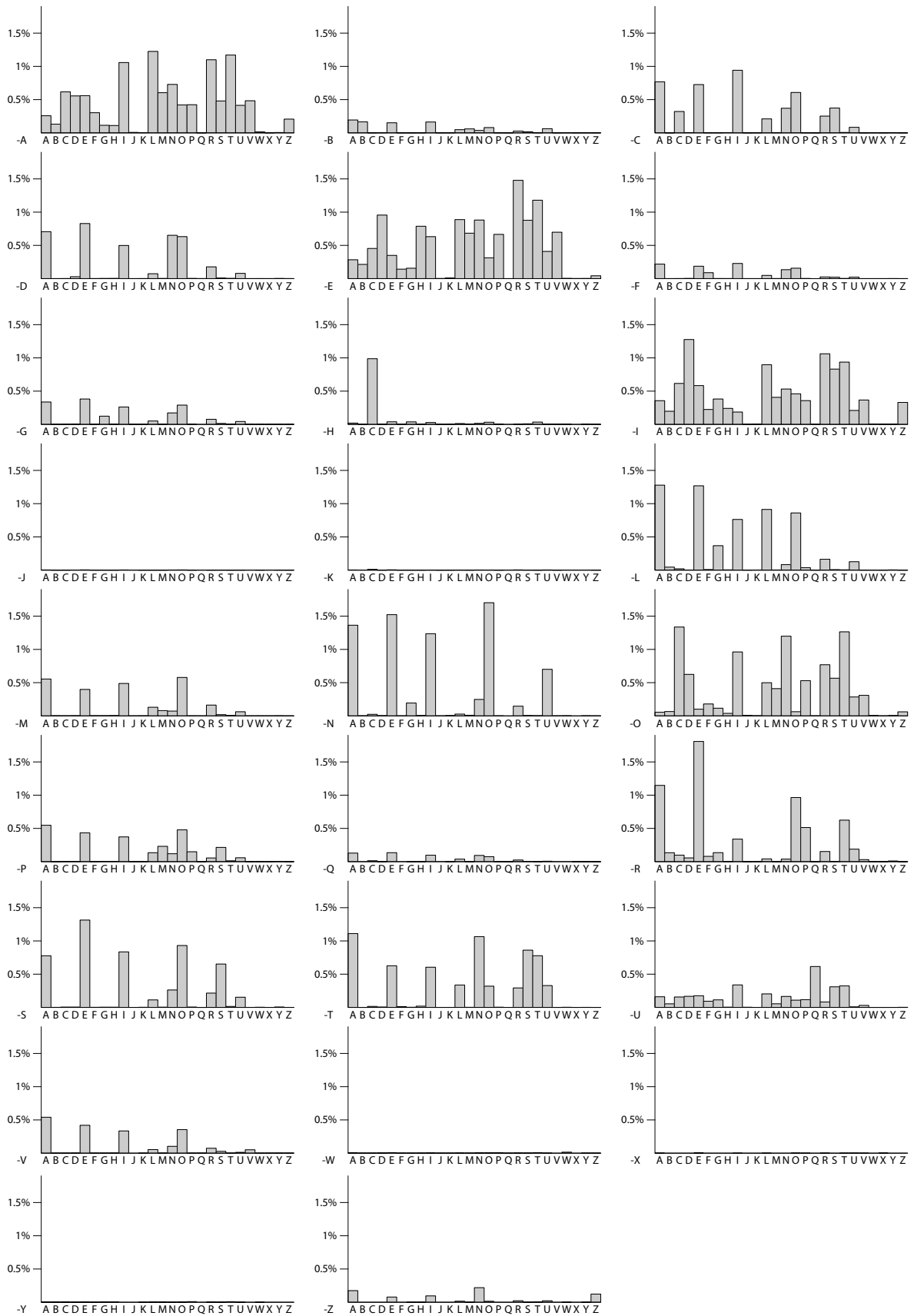
Per prima cosa è interessante visualizzare la frequenza delle doppie (barra a destra) confrontata con quelle delle stesse lettere singole (barra a sinistra). Per comodità abbiamo moltiplicato per dieci la scala delle doppie: ad esempio la «L» compare con una frequenza media del 6%, mentre la «LL» con una frequenza dello 0,9%. Da questa tabella possiamo riconoscere l'insieme {«L», «S», «T»}, anche se non siamo in grado di distinguere le tre lettere.



Nelle pagine seguenti vediamo la frequenza dei digrammi in cui la seconda lettera è fissata e quelli in cui la prima lettera è fissata. I più utili sono quello della combinazione «-H», cioè il digramma in cui l'«H» compare al secondo posto, che permette di riconoscere «C» ed «H»; e quello della combinazione «Q-», che permette di riconoscere «Q» ed «U».

Ora, se una delle lettere «L», «S» o «T» è seguita in percentuale significativa dalla «C», allora è sicuramente una «S».

A questo punto non rimane che provare a riconoscere delle lettere e tentare delle congetture, ricordando che più è breve il testo più alta sarà la variabilità statistica rispetto alla media.



Frequenza dei digrammi in italiano

