

## 5.1. Stabilisierungscharakter bei der Häufigkeitsanalyse

Max und Gloria erhalten drei verschiedene Texte von ihrem Lehrer, welche alle mit Caesar und dem gleichen Schlüssel chiffriert wurden.

- **Text 1:** 5 Buchstaben (datei1.txt)
- **Text 2:** 30 Buchstaben (datei2.txt)
- **Text 3:** 1 000 Buchstaben (datei3.txt)
- **Text 4:** 10 000 Buchstaben (datei4.txt)

Beide versuchen den passenden Schlüssel durch Häufigkeitsanalyse bei den unterschiedlichen Textlängen herauszufinden. Am Ende kommt Max zum Schluss:

*„Bei Text1 und Text2 war es schwieriger den Schlüssel zu finden, während bei Text3 bzw. Text4 der Schlüssel relativ schnell klar war“*

Auftrag:

- Wie kommt Max zu dieser Aussage?
- Vergleiche die relative Häufigkeit des am häufigsten vorkommenden Buchstabens - mit der zu erwartenden Häufigkeit des Buchstaben e: 17.4%. Was fällt dir auf?

**Bsp.** Der am häufigsten vorkommende Buchstabe d kommt 5-mal im Text vor, der Text hat insgesamt 30 Buchstaben, d.h. der Buchstabe d hat eine relative Häufigkeit von

$$\frac{\text{Anzahl des Buchstaben } d}{\text{Anzahl der Buchstaben im Text}} = \frac{5}{30} = 0.166666 \dots \approx 16.6\%$$

Anschließend vergleicht man den ermittelten Wert  $\frac{5}{30}$  mit dem zu erwarteten Wert 17.4%.

- Fülle nachfolgende Liste aus

Text	Text 1	Text 2	Text 3	Text 4
der(die) am häufigsten vorkommende(n) Buchstabe(n)				
Anzahl				
Relative Häufigkeit (in Prozent)				

- Warum fällt es ihm leichter bei Text3 bzw. Text4 einen Schlüssel zu finden als bei Text1 bzw. Text2?