

# UNI-Camp 2020

Language Science and Technology

Liebe UniCamp Teilnehmerinnen,

Schön, dass Ihr dabei seid! Da wir aufgrund der aktuellen Corona Situation ein etwas anderes Programm mit Euch durchführen, sind wir auf Eure Mitarbeit angewiesen. Im Folgenden findet Ihr einige Anregungen, was Ihr alles mit einem mehrsprachigen Korpus machen könnt. Denkt dran, Ihr seid nicht in der Schule und es ist gar nicht schlimm, wenn etwas nicht klappt! Pauline und Yasmin sind da um Euch zu helfen. Die Idee ist, dass die folgenden Abfragen nacheinander gemacht werden, da manches aufeinander aufbaut. Ihr müsst Euch aber nicht unbedingt an die Aufgabenstellungen halten, ihr dürft auch kreativ werden und eigene Abfragen an das Harry Potter Korpus stellen. Ihr könnt immer in der [Präsentation](#) nachschauen, wie man bestimmte Abfragen stellt. Eure Login-Daten für [CQP-Web](#) findet ihr vor Euch auf dem Tisch. Das Passwort ist für alle **12345!**

Tobt Euch aus und habt vor allem SPAB!!!

Pauline, Stefania und Yasmin

Fangt am besten mit dem [deutschen](#) Harry Potter Korpus an. Hier könnt ihr eure ersten eigenen Abfragen (Queries) ausprobieren:

# Das Harry Potter Corpus

## 1. Lemma oder Word ?

Grundform der Query: [lemma="einWort"] oder [word="einWort"]

Setze irgendein Wort in die Anführungsstriche. Was ist der Unterschied zwischen *word* und *lemma*?

Warum ist es nützlich diese Unterscheidung zu haben?

## 2. Buchstaben(folgen) festlegen

Wie kann man alle Wörter finden, die mit K (**oder** k) beginnen?

Wie kann man Wörter finden, die auf „ei“ enden?

Mit welcher Query kann ich alle Wörter, die „Zauber-/zauber“ **beinhalten** abfragen?

**Tipp**

Man kann hier reguläre Ausdrücke benutzen. Schau noch einmal in die Präsentation 😊!

### 3. Wortarten abfragen (pos)

Erinnerung: pos steht für Wortart. Jede Wortart hat ihr eigenes Kürzel. Dieses könnt ihr in den jeweiligen **Listen (s. unten)** nachschauen

- Grundform der Query: [pos="irgendeinPOSTag"]
- Setze irgendein pos-tag aus der Liste (für Deutsch, z.B. ADJA) ein und schau, was herauskommt.
- Wie kann man **Nomen** finden? Wie kann man **Eigennamen** finden?

### 4. Kombinierte Abfragen:

- Wie kann man Nomen **und** Eigennamen abfragen (es gibt mehrere Möglichkeiten)?
- Wie kann man nur Nomen finden, die mit P beginnen?  
Bsp.: Ich suche ein Verb, das mit f beginnt: [pos="V.\*" & word="f.\*"]
- Wie kann man **alle** Arten von Verben finden?
- Wie wird Harry beschrieben? Suche nach Adjektiv + Harry (hier musst du zwei aufeinanderfolgende Wörter abfragen. Tipp: [] [])
- Das gleiche kannst du auch für alle anderen Charaktere ausprobieren.

**Tipp**

Auch hier kann man reguläre Ausdrücke benutzen. Der Punkt (.) steht z.B. für einen weiteren Buchstaben, das Sternchen ( \* ) steht für beliebig viele Buchstaben, usw. Schau noch einmal in die Präsentation ☺!

Listen der POS-tags in den anderen Sprachen:

**Deutsch**

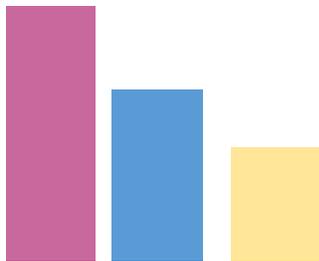
**Englisch**    **Latein**

**Spanisch**    **Franzoesisch**

Die POS-tags gelten nur im Corpus der jeweiligen Sprache!!!

## 5. Frequency Breakdown

- Welche drei **Namen** fallen am häufigsten im ersten Band ([deutsch](#))? Und im [englischen](#) Original? Hierfür müsst Ihr erst nach Namen suchen und dann einen Frequency Breakdown machen (s. Präsentation).
- Welche Nomen, Adjektive, Verben sind die häufigsten? Für Verben nehmt Ihr am besten nur die Vollverben ( [pos="VV.\*"] ), die sind interessanter als Hilfs- und Modalverben 😊



Für diese Abfrage müsst ihr in das Korpus der jeweiligen Sprache gehen:

[Englisch](#)

[Französisch](#)

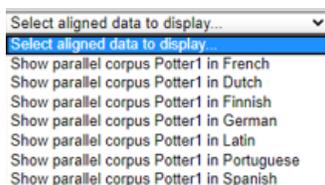
[Spanisch](#)

[Latein](#)

## 6. Übersetzung

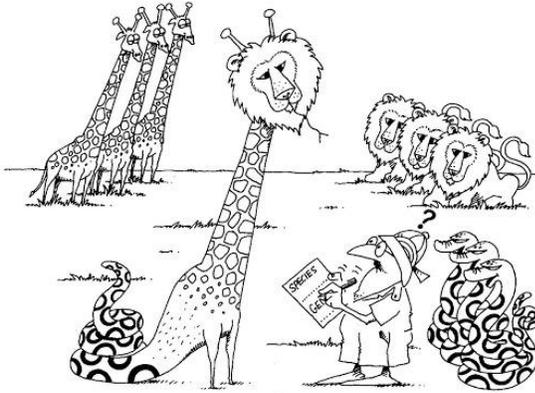
Die Übersetzung ins Französische, Spanische oder Latein kann nur vom [englischen Originalkorpus](#) aus angezeigt werden! Vom deutschen Korpus aus kann man sich auch das englische Original anzeigen lassen.

- Findet heraus, was Zauberstab, Zaubertrank, Zauberspruch auf Englisch, Spanisch oder Französisch heißt. Ihr könnt alle in einer Query abfragen. Es geht aber auch einzeln.
- Im Englischen ([Potter1 English](#)): Wann benutzt man *tall*, *high*, *big* und wann *small*, *short*, *little*? Wie wird es ins Deutsche übersetzt? Frage alle drei Wörter in **einer** Query ab. Lasse dir die alignierten Übersetzungen anzeigen!



(Klicke "aligned data to display" → Switch)

## 7. Kategorisierung



Hier könnt ihr eigene Annotationen in das Korpus einfügen, um euch eigene Fragen zu beantworten. Z.B. könntet ihr Personen in männliche und weibliche Charaktere einteilen, Adjektive in Kategorien einteilen (Aussehen, Charaktereigenschaften, etc.), Verben einteilen (Bewegung, Gefühl, etc.) usw. und hinterher Häufigkeiten bestimmen. Schaut euch dafür das folgende [Video](#) von Stefania an!