**Korpuslinguistik – Eine Einführung**

**Anwendungsbeispiel:**

Das Jugendwort des Jahres 2021 ist "Cringe" (Voss 2021). Sucht via Suchfunktion in dem von euch am häufigsten verwendeten Messenger-Dienst nach dem Wort "Cringe" (oder auch andere Wörter, die ihr häufig verwendet).

* Notiert euch in welcher Umgebung das Wort vorkommt – also welche anderen Wörter tauchen vor oder nach dem Wort "Cringe" auf?
* Notiert euch auch welche Worte am häufigsten zusammen mit "Cringe" vorkommen. Erstellt eine kurze Rangliste. Haltet eure Ergebnisse auf Padlet (www.padlet.com) fest oder alternativ könnt ihr auch das Arbeitsblatt mit der Tabelle benutzen.

**Was ist Linguistik?**

Linguistik untersucht "Sprache mit ihren aktuellen und historischen Formen und Verwendungsweisen" (Linguistik). In diesen Bereich fallen auch Fragestellungen zum "Verhältnis von Sprache, Kultur und Gesellschaft" (Linguistik). Zur Bearbeitung dieser Fragestellungen gibt es unterschiedliche Methoden. Die Korpuslinguistik ist eine dieser Methoden.

**Was ist Korpuslinguistik?**

Die Korpuslinguistik ist eine **wissenschaftliche Methode** die zur Beantwortung von bestimmten Forschungsfragen genutzt werden kann (Hirschmann, 2019, S. 1). Die Korpuslinguistik ist aber nicht für die Beantwortung aller Forschungsfragen geeignet, da die eingesetzten Methoden immer von den Forschungsfragen abhängig sind. Das folgende Zitat gibt eine breite Definition von Korpuslinguistik.

«Die ***Korpuslinguistik*** ist eine **empirische Methode** mit dem Ziel, **linguistische Forschungsfragen** zu bearbeiten. Eine empirische Methode zeichnet sich durch **bestimmte Datentypen** aus, die gezielt durch **experimentelle oder nicht-experimentelle Datenerhebungen** gewonnen werden» (Hirschmann, 2019, S. 1)

Bei der Korpuslinguistik handelt es sich also um die wissenschaftliche Auseinandersetzung mit einem **Korpus**. Das Korpus wird folgendermassen definiert:

««Ein ***Korpus*** ist eine **Sammlung von Textdaten**, also Sprache im Kontext, die dem **Zweck der linguistischen Auswertung** dient und eine […] Auswertung von […] **sprachlichen Merkmalen** zulässt.» (Hirschmann, 2019, S. 2)

Das ist ebenfalls eine sehr breite Definition. Im Grunde genommen ist ein Korpus ein Datensatz mit dem wissenschaftliche Analysen durchführen kann. Konkrete Beispiele dafür sind z.B. die ICE-Corpora der Englischen Sprachwissenschaft (International Corpus of English, 2021). Die Abkürzung ICE steht für International Corpus of English (International Corpus of English, 2021). Dabei handelt es sich um eine Sammlung von mehreren Korpora. Jeder dieser Korpora deckt eine Varietät des Englischen ab. Es gibt also, z.B. ein Korpus für Britisches, Amerikanisches, Kanadisches, Neuseeländisches, Jamaikanisches Englisch, usw. (International Corpus of English, 2021). Diese Korpora sind für den Vergleich der verschiedenen Englisch Varietäten geschaffen. Es gibt auch verschiedene Korpora in der Germanistik. Zum Beispiel gibt es die Plattform des digitalen Wörterbuch der deutschen Sprache (DWDS, 2021).

**Was ist das DWDS?**

Das Digitale Wörterbuch der deutschen Sprache (DWDS) ist unter dem Link <http://www.dwds.de> grösstenteils frei zugänglich. Das Ziel hinter dem DWDS war die Erstellung "eines umfassenden, jedem Benutzer über das Internet zugänglichen Wortinformationssystems, das Auskunft über den deutschen Wortschatz in Vergangenheit und Gegenwart gibt" (Geyken et al. 2017, S. 328). Zusätzlich zu der Wörterbuchfunktion des DWDS stellt die Plattform auch die Korpora, auf denen die Wörterbucheinträge basieren, zur Verfügung (Geyken et al. 2017, S. 328). Somit sind die Daten auf denen die Einträge basieren auch zum grössten Teil frei zugänglich (Geyken et al. 2017, S. 328). Die Plattform des DWDS ermöglicht "die gleichzeitige Suche in verschiedenen digitalen Wörterbüchern und Korpora sowie die Nutzung statistischer Analysewerkzeuge, die auf den Korpora basieren" (Bartz & Radtke 2014, S. 131).

**Was lässt sich mittels DWDS herausfinden?**

**Wortverlaufskurven:**  Zeigen die Verwendungshäufigkeit eines Wortes inner-

halb eines bestimmen Zeitabschnitts an.

**Wortprofile:** Zeigen welche Wörter in Verbindung mit einem spezif-

isch ausgewählten Stichwort vorkommen.

**DiaCollo-Kollokationsanalysen:** Zeigen typische Wortverbindungen zu einem Stichwort in

einem bestimmten Zeitabschnitt an.

**Literatur**

Bartz, Thomas und Nadja Radtke (2014). Digitale Korpora im Deutschunterricht: Didaktisches

Potenzial. In: ZGL 42/01, 130-143.

DWDS = Das Digitale Wörterbuch der deutschen Sprache. Hrsg. von der Berlin-Brandenburg-

ischen Akademie der Wissenschaften 2019. Online unter: https://www.dwds.de/wb/. <13.12.2021>.

Geyken, Alexander, Adrien Barbaresi, Jörg Didakowski, Bryan Jurish, Frank Wiegand u. Loth-

ar Lemnitzer (2017): «Die Korpusplattform des „DigitalenWörterbuchs der deutschen Sprache“ (DWDS)». In: ZGL 42/02, 327–344.

Hirschmann, Hagen: Korpuslinguistik. Eine Einführung. Berlin. 2019.

(o.A.). Linguistik. Hrsg. von der Universität Zürich. Online unter: https://www.ds.uzh.ch/de/lin

guistik.html. <13.12.2021>.

(o.A.). International Corpus of English. Hrsg. von der Universität Zürich Online unter: https://w

ww.ice-corpora.uzh.ch/en.html. <02.12.2021>.

Voss, Pauline (2021): Fremdscham liegt im Trend: «Cringe» ist das Jugendwort des Jahres

2021, In: NZZ. Online unter: https://www.nzz.ch/panorama/fremdscham-liegt-im-trend

t-cringe-ist-das-jugendwort-2021-ld.1651937. <02.12.2021>.

**Weiterführende Literatur**

Bubenhofer, Noah (2006–2019): Einführung in die Korpuslinguistik. Praktische Grundlagen

und Werkzeuge. Online unter: http://www.bubenhofer.com/korpuslinguistik/. <13.12.2021>.