ULDaF KARL-FRANZENS-UNIVERSITÄT GRAZ

**Korpusanalyse (Káňa)**

**Die wichtigsten korpuslinguistischen Begriffe**

|  |  |  |
| --- | --- | --- |
| **Begriff** | Erklärung | *Beispiel* |
| **Korpusmanager**  | etwa wie Suchmaschine im Internet  | *CQP, (Corpus Query Processor), Cosmas II, Bonito…* |
| **Suchfeld** | Abfragefenster |  |
| **Suchanfrage:** |  |  |
| ***Word*** | Wortform: beliebige Buchstaben- / Graphemhäufung | *saß, Ach, Ring, Tohuwabohu, lepschi* |
| ***Intervall*** | Wortteil, Affix (Vorsilbe/Nachsilbe) | Suche nach: *~saß, blöd~, ~ring~, ~erl* |
| ***Wortverbindung*** | mehrere Wörter | *an der Quelle, taube Nuss* |
| ***Lemma***  | Grundform eines Lexems; Lexikoneintrag (Nominativ Singular, bzw. Infinitiv Präsens) | *Apfel* (zu *Apfels, Äpfel, Äpfeln*); *hören* (zu *höre, höret, hörend, gehört*…) |
| ***Tag***  | Morphologische Kategorie (falls das Korpus getaggt ist); in jedem Korpus andere Kategorien (je nach Sprache) und anderes System (Manual notwendig!) | Suche nach allen Interjektionen: *ITJ* (InterCorp)Suche nach allen Imperativen: *VAIMP*(InterCorp) |
| **Suchoperatoren** | Wörter oder Zeichen für kompliziertere Abfragen: z.B. beim suchen des Verbs *einigen*; nicht das Adjektiv *einig(en)* (Manual notwendig!) | *nicht, oder,…**&, |* |
| **Platzhalter** | Zeichen zum Suchen von geleichzeitig mehreren Formen, die einen gemeinsamen Teil haben – z.B. alle Wörter, die auf *~erl* enden. (Manual notwendig!) | *.\*erl*(InterCorp)*++erl*(DeReKo) |
| **Token**  | laufende Wortform, alle vom Computer analysierten sprachlichen Zeichen (de Saussure: *„Wort ist ein Zeichen“*) | *oder so was , ich sitze an der Quelle . “* (11 tokens) |

|  |  |  |
| --- | --- | --- |
| **Konkordanz / Konkordanzzeile**  | Derjenige Teil des Textes, in dem das Gesuchte vorkommt | *besteht darin , das Problem* ***an der Quelle*** *in Angriff zu nehmen**kann man das Gas* ***an der Quelle*** *oder an der Küste stoppen.**Er machte* ***an der Quelle halt****, um Wasser zu trinken .* |
| **KWIC**  | „key word in context“: das Gesuchte in seinem (natürlichen) Umfeld (hier in ***Fettschrift***) | *Problem* ***an der Quelle*** *in* *machte* ***an der Quelle*** *halt**ich sitze* ***an der Quelle****. “* |
| **Beleg/Treffer**  | Ergebnis einer Korpussuche, belegt die (Nicht-)Existenz des Gesuchten im Korpus.  | (Eine Interpretation ist notwendig! Hier beginnt die intellektuelle Arbeit mit der Sprache.) |
| **Parallele(n)** | Einsprachige(r) Teil(e) in einem Parallelkorpus  | *de - en ‑ it* |
| **Segment**  | Sich in verschiedenen Sprachen entsprechende Passagen, die die durch ein Alignment entstanden sind. | *»Wir sind nämlich nicht blöd!«*  | *"'Cos we're not stupid!"* |
| **Alignment** | Anordnung der sich entsprechenden Textpassagen im Parallelkorpus. | *Sicherheitsmaßnahmen hingegen, die die Freiheitsrechte beschneiden, sind Maßnahmen, die dem Gesetz die Sicherheit entziehen. Somit sind sie die Quelle von Unsicherheit und Barbarei.* | *However, security measures that diminish the guarantees of freedom are measures that remove certainty from the law , and are thus the source of insecurity and barbarities.* |
| **Frequenz / Häufigkeit**  | Absolute oder relative Häufigkeit eines Wortes, Lexems etc. im Korpus |  |
| **Kookkurrenz / Kollokation**  | semantische oder grammatikalische Verbindungen | *eine/keine seriöse Quelle**von der Quelle bis zur Mündung**Quelle angeben* |
| **Mi-Score/ T-Score** | „mutual information“ Wahrscheinlichkeit der Kookkurrenz / „test score“ relativiert Mi-Score |  |
| **LLR** | „log-likelyhood-ratio“ Wahrscheinlichkeitsquotient der Kookkurrenz | je höher der Wert, desto sicherer das Vorkommen der Kollokatoren im selben Kontext |
| **corpus-based / korpusbasiert**  | verifiziert bisherige Hypothesen über Sprache(n) | Ist das Wort *Schmankerl* tatsächlich nur ein Austriazismus? |
| **corpus-driven / korpusgesteuert**  | betrachtet Korpus als „Ausgangspunkt“ für die Beschreibung der Sprache | Was kommt rechts vom Wort *Schmankerl* vor? |

Mehr z.B. in Káňa, T: *Sprachkorpora in Unterricht und Forschung DaF/DaZ*. Brno 2014. Online: <https://is.muni.cz/repo/1201835/cs?lang=en>