**Die wichtigsten korpuslinguistischen Begriffe**

|  |  |  |
| --- | --- | --- |
| **Begriff** | Erklärung | *Beispiel* |
| **Korpusmanager**  | etwa wie Suchmaschine im Internet  | *CQP, (Corpus Query Processor), Cosmas, Bonito…* |
| **Suchfeld** | Abfragefenster |  |
| **Suchanfrage:** |  |  |
| ***Word*** | Wortform: beliebige Buchstaben/ Graphemhäufung | *saß, Ach, Ring, usw,*  |
| ***Interval*** | 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. alles Wörter, die auf *~erl* enden. (Manual notwendig!) | *.\*erl*(InterCorp)*++erl*(DeReKo) |
| **Token**  | laufende „Wort“form, alle vom Computer analysierten sprachliche Zeichen (*„Wort ist ein Zeichen“* de Saussure) | *oder so was , ich sitze an der Quelle . “* (11 tokens) |
| **Konkordanz/Konkordanzzeile**  | Derjenige Teil des Textes, in dem das Gesuchte vorkommt |  |
| **KWIC**  | „key word in context“: das Gesuchte in seinem (natürlichen) Umfeld | A98 Auslöser war eine einseitige **Sprachenpolitik** und Unterdrückung derE98 etwa durch die einseitige **Sprachenpolitik** und die Arbeitsplatz-N98 Bedeutung einer systematischen **Sprachenpolitik** und die Förderung derP93 sein, gilt es umzudenken, was **Sprachenpolitik** und -unterricht |
| **Beleg/Treffer**  | Ergebnis einer Korpussuche, belegt die (Nicht-)Existenz des Gesuchten im Korpus. (Interpretation notwendig!) |  |
|  |  |  |
| **Parallelen** | Einsprachige Teile in einem Parallelkorpus  | *de – en - it* |
| **Segment**  | Entsprechende Passage in einer anderen Sprache.Entsprechung, die durch Alignment entstanden ist. | *» Wir sind nämlich nicht blöd ! «*  | *" 'Cos we 're not stupid ! "* |
| **Alignment** | Anordnung der sich entsprechenden Textpassagen im Parallelkorpus. |  |
| **Kookkurrenz/ Kollokation**  | semantische oder grammatikalische Verbindungen |  |
| **Frequenz/ Häufigkeit**  | Absolute oder relative Häufigkeit eines Wortes, Lexems etc. im Korpus (statistische Angabe) |  |
| **Mi-Score/ T-Score** | „mutual information“(statistische Angabe) |  |
|  |  |  |
| **LLR** | „log-likelyhood-ratio“ (statistische Angabe) |  |
| **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? |