

KORPUSANALYSE (Káña)

DIE WICHTIGSTEN KORPUSLINGUISTISCHEN BEGRIFFE

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

Konkordanz / Konkordanzzeile	Derjenige Teil des Textes, in dem das Gesuchte vorkommt	<i>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 .</i>	
KWIC	„key word in context“: das Gesuchte in seinem (natürlichen) Umfeld (hier in <i>Fettschrift</i>)	<i>Problem an der Quelle in machte an der Quelle halt ich sitze an der Quelle. “</i>	
Beleg/Treffer	Ergebnis einer Korpusuche, 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	<i>de - en - it</i>	
Segment	Sich in verschiedenen Sprachen entsprechende Passagen, die die durch ein Alignment entstanden sind.	<i>»Wir sind nämlich nicht blöd!«</i>	<i>”Cos we're not stupid!”</i>
Alignment	Anordnung der sich entsprechenden Textpassagen im Parallelkorpus.	<i>Sicherheitsmaßnahmen hingegen, die die Freiheitsrechte beschneiden, sind Maßnahmen, die dem Gesetz die Sicherheit entziehen. Somit sind sie die Quelle von Unsicherheit und Barbarei.</i>	<i>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.</i>
Frequenz / Häufigkeit	Absolute oder relative Häufigkeit eines Wortes, Lexems etc. im Korpus		
Kookkurrenz / Kollokation	semantische oder grammatikalische Verbindungen	<i>eine/keine seriöse Quelle von der Quelle bis zur Mündung Quelle angeben</i>	
Mi-Score/ T-Score	„mutual information“ Wahrscheinlichkeit der Kookkurrenz / „test score“ relativiert Mi-Score		
LLR	„log-likelihood-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 <i>Schmankerl</i> tatsächlich nur ein Austriazismus?	
corpus-driven / korpusgesteuert	betrachtet Korpus als „Ausgangspunkt“ für die Beschreibung der Sprache	Was kommt rechts vom Wort <i>Schmankerl</i> 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>