

Il Corpus: Terminologia utile

Italiano	Ceco
corpus / subcorpus	corpus / subcorpus
ricerca per lemma	vyhledávání podle lemmat
Lemma (per esempio "bello" per "bella", "belli" "belle")	lemma: je základní tvar pro nějaký výraz
word	word – slovní tvar – konkrétní tvar či tvar interpunkčního znaménka
concordanza = Ricerca dei contesti all'interno dei quali occorre una certa espressione linguistica	konkordance= seznamu všech konkordančních řádků, tedy všech výskytů hledaného výrazu (seznamu všech výskytů hledaného slova nebo jevu)
posizione (riguarda la concordanza, se voglio vedere cosa precede o segue la parola data KWIK)	pozice
KWIK = key word in context	KWIK
Collocazione : combinazioni fraseologiche	Kolokace : např. slovní spojení
token = parola grafica, costituente grafico, preceduto e seguito da spazio bianco (quindi parole, segni di punteggiatura, numeri ecc)	Token
Tag tagset = serie di etichette morfosintattiche applicate dal software specifico (PoStagger)	tag= sumarizuje gramatickou informaci o hledaném slově v konkrétním kontextu (tagset: per il ceco: 16 pozicí – viz PoS)
PoS = part of speech	Např. N:noun, ADJ, adjective, ADV adverb, PRON pronoun atd.
markup = etichetta che si dà a una sequenza di caratteri (parola o gruppo di parole) e che ne definisce il valore. Per l'Annotazione linguistica = esplicitazione dei valori linguistici rispetto ai livelli dell'analisi: sintattica, semantica, morfologica...)	markup
occorrenza = distingue le singole parole all'interno della sequenza testuale	<i>frekvence</i>