Oefeningen OpenSoNaR tutorial

9 oktober 2020

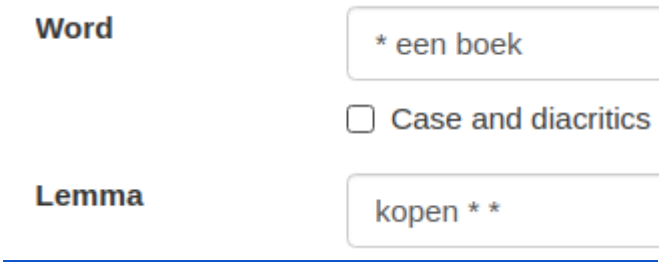
Simple Search

1. Zoek op voorkomens van *pannenkoek* met en zonder tussen-n. Hoe vaak komen beide vormen in het corpus voor?
2. Odijk (2020) stelt dat de combinatie *heel af en toe* welgevormd is, maar *erg af en toe* niet. Wat vinden we in het OpenSoNaR voor deze expressies?
3. Corpora bevatten teksten, geschreven/geproduceerd door mensen. Mensen maken fouten en die fouten zitten dus ook in het corpusmateriaal. Een voorbeeld hiervan is *episch centrum* in plaats van *epicentrum*. Hoe vaak komt dit in het corpus voor? In welk teksttype komt dit vooral voor?

Extended Search

4. Zoek het woord *graven* op in het CGN. Groepeer op lemma. Welke lemmata zie je en wat valt je daaraan op?
5. Is er verschil tussen Nederland en Vlaanderen in het gebruik van deze woorden: *verdiep* versus *verdieping*? Onderzoek dit voor het CGN.
6. Morfologische woordstructuur is niet geannoteerd in OpenSoNaR. Je kunt daar dus ook niet op zoeken. Met wildcards of reguliere expressies kun je dit gemis gedeeltelijk opvangen. Zoek in extended binnen het CGN naar woorden die dezelfde morfologische structuur hebben als *betrokkenheid*.
7. Zoek uit welke varianten van de uitdrukking *te allen tijde* voorkomen in het corpus.
8. In woorden als *lopen* kan de slot-n wel of niet worden uitgesproken. Zoek met behulp van de fonetische transcriptie (veld *phonetic*) uit welke variant het meest karakteristiek voor welk land is. Hint: kijk hoe de transcriptie werkt door eerst het woord op te zoeken en dan te groeperen op phonetic.
9. In extended search kan je niet alleen woorden invoeren maar ook simpele zoekpatronen. Zo kan je “|” gebruiken voor meerdere mogelijkheden, en kan je meerwoordpatronen als “* boek” invullen, waarbij de wildcard * voor een willekeurig woord staat. Daarbij kan je ook

meerdere velden combineren, zoals in



Zoek in 1 zoekopdracht, met groepering, uit wat de relatieve frequenties van de hulpwerkwoorden *hebben* en *zijn* direct voor het deelwoord *gefietst* zijn.

Advanced search

10. Van Noord en Odijk (2016) beweren dat *'m* als onderwerp wel in Vlaanderen maar niet in Nederland gebruikt wordt. Verder stellen ze dat in Nederland in plaats daarvan *ie* veel meer gebruikt wordt. Kan deze claim onderbouwd worden door het CGN-corpus? Gebruik de opeenvolging onderschikkend voegwoord + *'m / ie* als benadering voor *'m / ie* als onderwerp.
11. Voorzetseluitdrukkingen. Zoek naar uitdrukkingen zoals *in tegenstelling tot*, d.w.z van de vorm voorzetsel + enkelvoudig zelfstandig naamwoord + voorzetsel. Groepeer op lemma en geef de top 5.
12. Welke zelfstandig naamwoorden volgen vaak op het lemma *licht*?
13. Onderzoek de frequenties van de "rode" en de "groene" volgorde van de werkwoordelijke groep in het CGN. Toelichting:
    i. Groene volgorde: deelwoord + persoonsvorm *(gelopen heb*)
    ii. Rode volgorde: persoonsvorm + deelwoord (*heb gelopen*)

Expert search

14. Hoeveel verschillende spellingen van het woord *online* kom je in het corpus tegen? Houd rekening met koppeltekens en spaties.
15. Uitdrukkingen zoals *straat in straat uit*, d.w.z. zelfstandig naamwoord + achterzetsel + zelfstandig naamwoord + achterzetsel,waarbij de twee zelfstandige naamwoorden identiek zijn, gegroepeerd op woord. Zoek zulke uitdrukkingen en geef de top 5.
16. Scheidbare werkwoorden zijn niet als zodanig geannoteerd in OpenSoNaR. Als je dus zoekt op lemma *aanklampen,* vind je niet de volgorde "zij klampte hem aan". Probeer met een CQL-patroon (eventueel gevolgd door groepering) uit te zoeken welke scheidbare werkwoorden met werkwoordelijk deel *lichten* in deze volgorde (werkwoordelijk deel gaat vooraf aan partikel) voorkomen in het corpus.