

4. Vilka ord är särskilt framträdande i en text eller samling av texter?

Nyckelord kan innebära flera saker. I ett språkteknologiskt sammanhang är nyckelord sådana ord som har större förekomst (eller högre frekvens) än förväntat i en text eller samling texter. Ofta är det ord som kan vara förväntade, exempelvis ord som handlar om ekonomi i en budgetplanering eller i andra fackspråksbetonade texter, men inte alltid. Att undersöka statistiskt belagda i texter eller andra språkmateriel är därför en användbar metod för både språkvetare och andra forskare.

Använd funktionen Keywords (bild 1 och 2) för att se vilka ord som har prominens i den valda korpusen svTenTen20.

- Vilka typer av ord förekommer? Går det att se mönster för exempelvis olika ordklasser eller olika typer av innehåll?
- Keywords ger även resultat för flerordsfraser. Bekanta dig även med denna typ av sökträffar och resultat. Vilka typer av mönster eller mer preciserade sökningar kan göras för flerordsfraser?
- Prova gärna Keywords för andra korpusar i Sketch Engine, både på svenska eller för andra språk.

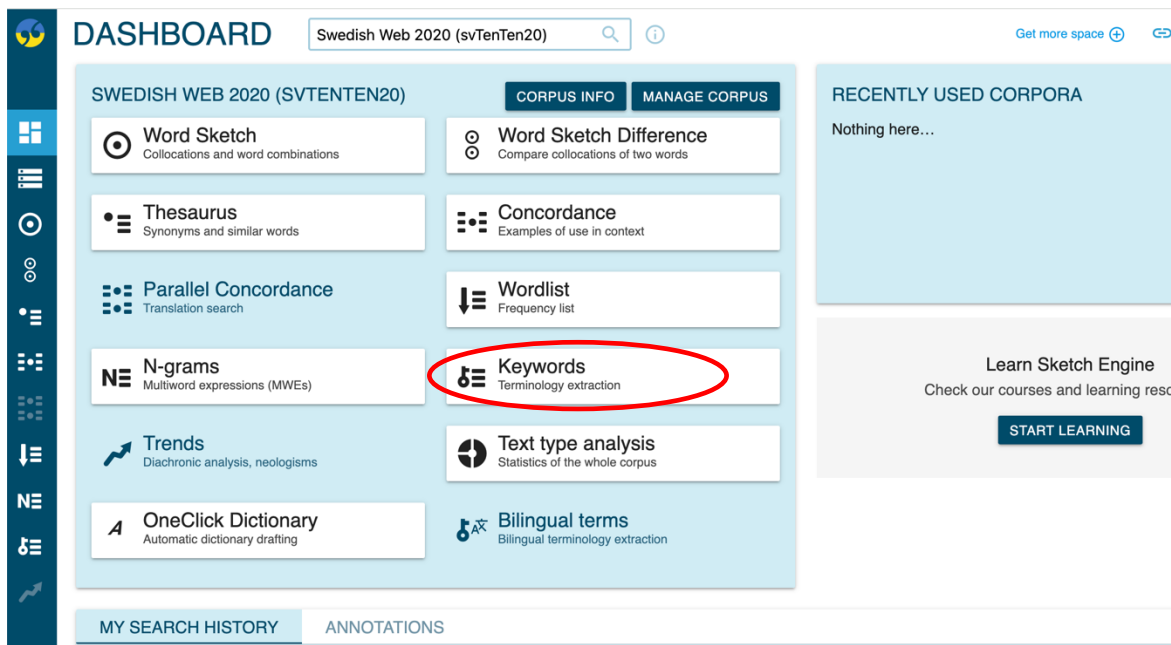


Bild 1: funktionen Keywords på Sketch Engines dashboard

KEYWORDS Swedish Web 2020 (svTenTen20) ⓘ

BASIC ADVANCED LEARN ⓘ

This tool compares corpora and identifies what is unique or typical. The selected corpus is compared to a reference corpus to identify these key data:

- Keywords** individual words (any token can be included)
- Terms** key multi-word expressions in a format typical of terminology in the language of the corpus
- N-grams** key multi-word expressions (any sequences of tokens). Only items which appear more frequently in the selected corpus than in the reference corpus are included. The results indicate what is typical of the selected corpus compared to the reference corpus.

Keyword and term extraction from a preloaded corpus
The extraction from a large preloaded corpus can take several minutes to complete.

I KNOW WHAT I'M DOING: GO

Bild 2: information om output för funktionen Keywords