

Task 1.

ICLE-FR, ICLE-DU, ICLE-SP, ICLE-TSW の4つのグループから into の頻度を検索し、10万語の相対頻度で比較しなさい。

【ポイント】

- コーパス・サイズの異なるグループの比較は相対頻度で行う
- パーセントでも統計処理上はよい場合もあるが、通例、コーパスの規模に合わせて100万語、10万語、5万語などで正規化 (normalize) することが多い

【基本ステップ】

- 1) それぞれサブコーパスを指定して1つずつ into を検索、粗頻度 (raw frequency) を記録する
- 2) 各サブコーパスの総語数を調べる
- 3) 総語数を10万語換算にして比較する

《注意点》

- Sketch Engine の「総語数」にはタグも句読点なども含まれている。
- サブコーパスを1つずつ検索してもよいが、Sketch Engine ではもう少し便利な検索手法がある

Task 2.

名詞+into, 自動詞+into, 他動詞+into を検索し、10万語の相対頻度で比較してみなさい。

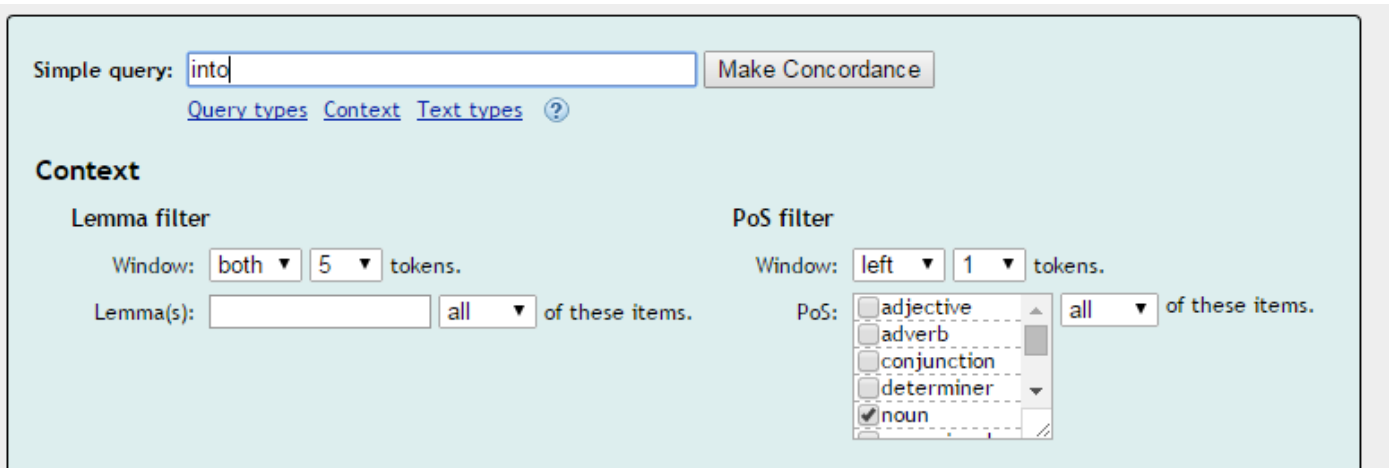
【ポイント】

- 「品詞」と単語を組み合わせた検索をする方法を学ぶ

【基本ステップ】

- 1) Simple query + POS filter を使用する方法

◎ 名詞+into



The screenshot shows the Sketch Engine search interface. At the top, there is a 'Simple query' field containing the word 'into' and a 'Make Concordance' button. Below the query field are links for 'Query types', 'Context', 'Text types', and a help icon. The 'Context' section is divided into two parts: 'Lemma filter' and 'PoS filter'. The 'Lemma filter' section has a 'Window' dropdown set to 'both' and a value of '5' tokens, and a 'Lemma(s)' field with 'all' selected. The 'PoS filter' section has a 'Window' dropdown set to 'left' and a value of '1' tokens, and a 'PoS' dropdown menu with 'noun' selected. The dropdown menu also shows other options: adjective, adverb, conjunction, and determiner.

《注意点》

- ただし V+NP+into との区別はつかない。コンコーダンスをよく観察してみよう。

2) CQL を使用する方法

正規表現ルール

[tag="NN|NNS"] [lemma="into"]

AB A または B

[tag="N.*"] [lemma="into"]

. : 何か 1 文字 * : その 0 回位所の繰返し

◎ 動詞+into

[tag="V.*"] [lemma="into"]

《注意点》

- これだと何がまずい??
- 自動詞を選ぶにはどうする??

[tag="V[BHV][^N]"] [lemma="into"]

[BHV] : B か H か V

[^N] : N 以外

[tag="V[BHV]([^N])?"] [lemma="into"]

()? : optional

◎ 動詞+名詞+into

[tag="V[HV][^N]"][1,3][tag="N.*"] [lemma="into"]

1 語から 3 語の間に
次のものがくる

Task 3.

「自動詞+into」を検索して、その自動詞の語彙的ヴァリエーションを頻度表で示しなさい

【ポイント】

- Task 2 の検索を行った後で、動詞だけリストにする方法・・・少し難しい

CQL: [tag="V[HV][^N]"] Default attribute: word

[Tagset summary](#)

Context

Lemma filter
Window: right 1 tokens.
Lemma(s): into all of these items.

PoS filter
Window: both 5 tokens.
PoS: adjective adverb conjunction determiner noun all of these items.