

## BNCweb (CQP): Simple Query Syntax

The simple query (SQ) feature of BNCweb (CQP edition) allows you to perform sophisticated corpus searches without learning a complex query language. Start with simple and intuitive word or lemma queries, then add more advanced features as you become familiar with the SQ syntax. If you want to use the expert CQP mode, look up your simple queries in the history and have them automatically translated into the full CQP syntax.

### Basic word form searches

- To search for word forms, simply type them into the query field.
- Use wildcards to generalize your searches: `?` (one arbitrary character), `*` (0 or more characters), `+` (1 or more characters)
- Protect wildcards and other meta-characters with backslash `\` to match literally, e.g. `\?` to search for a question mark. SQ syntax uses the following meta-characters which you may need to protect:

`? * + [ ] ( ) | _ { } < >`

(in certain contexts, other characters may also have special meaning)

- Specify comma-separated alternatives for substrings in square brackets, e.g. `[over,under]+ment` or `neighbo[u,r]`

### Matching parts-of-speech

- Search for word forms with specific POS tag: `lights_NN2`
- You can use wildcards in the POS constraint, too: `light+_N*`
- CQP allows non-lexical POS searches: `*_AJS` or simply `_AJS`

### Lemma queries

- Match lemma in curly braces: `{light}`, with wildcards: `{*ize}`, with syntactic categories: `{*ize/VERB}` or `{*ize/V}`, and even in combination with a POS tag `{light/N}_NP*`

- “POS lemmas” are abbreviations for major POS categories {N}, {V}, {A}, {Adv}, {Det}, {Prep}, {\$}, etc.: e.g. [under,over]+\_{N}

## Word sequences

- Type multiple words, separated by blanks, to search a specific expression in the corpus: e.g. **birds of a feather**
- NB: punctuation, possessives, contractions are separated (and some have to be escaped): search **they 've** and **wo n't he \?**
- Each word in a sequence can make full use of wildcards, POS constraints, lemma search, etc.: {bird/N} \_{Prep} \_{Det} \_{N}
- + stands for an arbitrary word, \* for an optional arbitrary word; ellipsis-like dots (.., ..., etc.) are a shorthand notation for such gaps

## Advanced word patterns

- Group word sequences with parentheses, separating alternatives within each group: **the (most \_AJ0 | \_AJS) \_{N}**
- Groups can be repeated by appending ? (0 or 1), \* (0 or more) or + (1 or more) to the closing bracket (no blanks allowed!):  
**{buy/V} ( \_{Det} )? ( \_{A} )\* \_{N}**

## Proximity queries

- Alternative query syntax allows window-based searches for word combinations, e.g. “unveil” accompanied by “statue” (within 5 words): **{unveil} <5> statue** (or <s> for sentence context)
- Matches of a proximity query include only the left element
- One-sided context: **good >3> bad** (“good” precedes “bad”) and **good <3< bad** (“bad” precedes “good”)
- Proximity queries can be nested with parentheses:  
**{unveil} <5> ( statue <2< bronze )**
- It is not possible to combine proximity queries with word patterns