

## Workshop: How to query corpora / Como pesquisar em corpora

### CWB - Sintaxe de pesquisa / Query syntax

O sistema IMS-CWB permite dois tipos de consulta:  
IMS-CWB allows for two types of query:

1. **Concordâncias:** pedido de frases no corpus que contemplem a palavra/expressão de pesquisa.  
**Concordance:** instances of the corpus featuring the queried word/expression.
2. **Distribuição:** pedido de quantificação/número de vezes que a palavra/expressão ocorre.  
**Distribution:** how often the queried word/expression is featured in the corpus.

### Alguns operadores de pesquisa / Some query operators

Op.	Ex.	Descrição / Description
"..."	"casa" "house"	As aspas balizam palavras e definem valores dos atributos. The quotation marks identify the borders of words and define values for the attributes.
[ ]	[lema="amar"] [lemma="love"]	Os parênteses rectos especificam os atributos das palavras/expressões. The square brackets specify the attributes of words/expressions.
( )	"(a de)" "gr(e a)y"	Agrupa lemas, palavras, expressões ou construções. Groups lemmas, words, expressions or constructions.
	"a de" "gray grey"	Operador de alternativa, disjunção. Operator for alternation, disjunction.
!	[word!="..."]	Operador de negação; excluir o assinalado. Operator for negation; excludes the text between quotation marks.
.	"pens." "car."	Representa qualquer carácter (apenas um); ex.: a palavra pode terminar em qualquer carácter: penso, pensa. Represents any character (only one); eg.: the word may end in any character: cart, card, etc.
*	"pens.*" "light.*"	Quantificador de caracteres: zero ou mais vezes; ex.: a palavra pode terminar por vários caracteres: penso, pensamento, pensamento, etc. Quantifier of characters: zero or more often; eg.: the word may end in a string of characters: lights, lighter, lightest, lighting, etc.
?	"cas.?" "arm.?"	Quantificador de caracteres: zero ou uma vez; ex.: casa, caso, etc. Quantifier of characters: zero or once; often; eg.: arm, arms, army, etc.
+	"cas.+" "arm.+"	Quantificador de caracteres: uma ou mais vezes; ex.: casa, casas, casamento, castigo, etc. Quantifier of characters: once or more often; eg.: army, armour, armchair, armpit, etc.
&	[word="casa" & pos="N"] [word="casa" & pos="N"]	Operador de conjunção: combina atributos para especificar elementos de pesquisa (no exemplo, a palavra casa está definida como 'nome'). Operator for conjunction: combines attributes to specify the queried item (in the example, the word house is defined as a 'noun')
=	[pos="N"]	Define o valor do atributo. Defines the value of the attribute.

CETEMPúblico	COMPARA	OPUS
<a href="http://www.linguateca.pt/cetempublico/">http://www.linguateca.pt/cetempublico/</a>	<a href="http://www.linguateca.pt/COMPARA/">http://www.linguateca.pt/COMPARA/</a>	<a href="http://opus.lingfil.uu.se/bin/opuscqp.pl">http://opus.lingfil.uu.se/bin/opuscqp.pl</a>
<b>Atributos/Attributes</b>	<b>Atributos/Attributes</b>	<b>Atributos/Attributes</b>
[word="..."] forma base / base form [lema="..."] palavra que corresponde à 'unidade lexicográfica' / form base corresponding to the 'lexicographic unit' [pos="..."] categoria gramatical / part-of-speech tags [temcagr="..."] tempo e modo verbal; caso pronominal e graus dos adjectivos / verb tense and mode, pronominal case and degrees of adjectives [pessnum="..."] pessoa e número / person and number [gen="..."] género / gender [func="..."] informação sintáctica / syntactic information	[word="..."] forma base / base form [lema="..."] palavra que corresponde à 'unidade lexicográfica' / base form corresponding to the 'lexicographic unit' [pos="..."] categoria gramatical / part-of-speech tags [temcagr="..."] tempo e modo verbal; caso pronominal e graus dos adjectivos / verb tense and mode, pronominal case and degrees of adjectives [pessnum="..."] pessoa e número / person and number [gen="..."] género / gender [emp="..."] locuções / phrases	[word="..."] forma base / base form [lem="..."] palavra que corresponde à 'unidade lexicográfica' / base form corresponding to the 'lexicographic unit' [pos="..."] categoria gramatical / part-of-speech tags
<b>Alguns valores / Some values</b>	<b>Alguns valores / Some values</b>	<b>Alguns valores / Some values</b>
a) [pos="..."]  <b>Categoria gramatical / Part of speech</b>  ADJ - adjetivo / adjective ADV - advérbio / adverb DET - determinante (artigo, pronome ou adjetivo) / determiner (article, pronoun or adjective)	a) [pos="..."]  ADJ - adjetivo / adjective ADJn - adjetivo nominal / nominal adjective ADV - advérbio / adverb DET - determinante e pronome (menos os pessoais) / determiner and pronoun (except personal ones) DEArt - artigo definido / definite article	[pos="..."]  CC - conjunção de coordenação / coordinating conjunction CD - numeral cardinal / cardinal number DT - determinante / determiner EX - verbo <i>haver</i> / existencial there FW - palavra estrangeira / foreign word IN - preposição ou conjunção subordinativa / preposition or subordinating conjunction JJ - adjetivo / adjective JJR - adjetivo comparativo / adjective, comparative JJS - adjetivo superlativo / adjective, superlative MD - modo / modal N.* - qualquer nome comum ou próprio /

<p><b>DET_ardt</b> - artigo definido / definite article  <b>DET_arti</b> - artigo indefinido / indefinite article  <b>DET_dem</b> - demonstrativo / demonstrative  <b>DET_poss</b> - possessivo / possessive  <b>KS</b> - conjunção subordinativa / subordinate conjunction  <b>KC</b> - conjunção coordenativa / coordinating conjunction  <b>N</b> - nome comum / common noun  <b>PROP</b> - nome próprio / proper noun  <b>NUM</b> - numeral / number  <b>PERS</b> - pronome pessoal / personal pronoun  <b>PRP</b> - preposição / preposition  <b>SPEC</b> - especificador – pronomes / specifier - pronouns  <b>SPEC_dem</b> - demonstrativo / demonstrative  <b>SPEC_rel</b> - relativo / relative  <b>SPEC_interr</b> - interrogativo / interrogative  <b>V</b> - verbo / verb</p> <p><b>Contrações / Contractions</b></p> <p><b>PERS\+PERS</b> - pronomes pessoais / personal pronouns  <b>PRP\+DET</b> - preposição e determinante / preposition plus determiner  <b>V\+PERS</b> - verbo e pronome pessoal / verb plus personal pronoun  <b>V_fmcl\+PERS\+PERS</b> - verbo de oração principal e pronomes / main clause verb plus personal pronouns</p> <p><b>b) [temcagr="..."]</b></p> <p><b>Tempo e modo verbal / Verb tense and mode</b></p> <p><b>PR_IND</b> - Presente do Indicativo / Present tense  <b>IMPF_IND</b> - Imperfeito do Indicativo / Past tense  <b>PS_IND</b> - Pretérito Perfeito do Indicativo / Past tense  <b>MQP_IND</b> - Mais-Que-Perfeito do Indicativo / Past tense  <b>FUT_IND</b> - Futuro do Indicativo / Future tense  <b>PR_SUBJ</b> - Presente do Conjuntivo / Present tense  <b>IMPF_SUBJ</b> - Imperfeito do Conjuntivo / Past tense  <b>FUT_SUBJ</b> - Futuro do Conjuntivo / Future tense  <b>IMP</b> - Presente do Imperativo / Present tense  <b>COND</b> - Condicional / Conditional tense  <b>INF</b> - Infinitivo / Infinitive  <b>GER</b> - Gerúndio / Gerund  <b>PCP</b> - Particípio / Past participle</p> <p><b>Casos pronominais / Pronominal cases</b></p> <p><b>NOM</b> - nominativo / nominative  <b>ACC</b> - acusativo / accusative  <b>DAT</b> - dativo / dative</p> <p><b>c) [pessnum="..."]</b></p> <p>número: <b>S</b> (singular)  / number <b>P</b> (plural)  <b>S P</b> (singular, plural)</p> <p>peessoa: <b>1S ; 2S ; 3S</b> (1<sup>a</sup>, 2<sup>a</sup>, 3<sup>a</sup> pessoa singular)  / person (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> person singular)  <b>1P ; 2P ; 3P</b> (1<sup>a</sup>, 2<sup>a</sup>, 3<sup>a</sup> pessoa plural)  (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup> person plural)  <b>1 3S</b> (1<sup>a</sup> e 3<sup>a</sup> pessoa do singular)  (1<sup>st</sup> and 3<sup>rd</sup> person singular)  <b>etc.</b></p> <p><b>d) [gen="..."]</b></p> <p><b>M</b> (masculino / masculine)  <b>F</b> (feminino / feminine)  <b>M F</b> (indeterminado / undetermined)</p>	<p><b>DETarti</b> - artigo indefinido / indefinite article  <b>DEtdem</b> - demonstrativo / demonstrative  <b>KC</b> - conjunção coordenativa / coordinating conjunction  <b>KS</b> - conjunção subordinativa / subordinate conjunction  <b>N</b> - nome comum / common noun  <b>Nprop</b> - nome comum iniciado por maiúscula/ capitalized proper noun  <b>PROP</b> - nome próprio / proper noun  <b>N_PROP</b> - nome comum que pode ser também nome próprio / common noun that can also be a proper noun  <b>NUM</b> - numeral  <b>NUMcard</b> - numeral cardinal / cardinal number  <b>PERS</b> - pronome pessoal / personal pronoun  <b>PERSrefl</b> - pronome reflexo / reflexive pronoun</p> <p><b>Contrações / Contractions</b></p> <p><b>PERS\+PERS</b> - pronomes pessoais / personal pronouns  <b>PERSrefl\+PERS</b> - pronome pessoal reflexo e pronome pessoal / reflexive pronoun plus personal pronoun  <b>PRP</b> - preposição / preposition</p> <p><b>PRP\+PERS</b> - preposição e pronome pessoal / preposition plus personal pronoun  <b>PRP\+PERSrefl</b> - preposição e pronome pessoal reflexo / preposition plus reflexive personal pronoun  <b>SPEC</b> - pronome que funciona como nome (menos os pessoais) / pronoun that serves as noun (except personal ones)  <b>SPECdem</b> - pronome demonstrativo / demonstrative pronouns  <b>SPECrel</b> - pronome relativo / relative pronoun  <b>V</b> - verbo / verb</p> <p><b>V\+PERS</b> - verbo e pronome pessoal / verb plus personal pronoun  <b>V\+PERSrefl</b> - verbo e pronome reflexo / verb plus reflexive pronoun  <b>V\+PERS\+PERS</b> - verbo e pronomes pessoais contraídos / verb plus contracted personal pronouns</p> <p><b>b) [temcagr="..."]</b>  ↓  (ver etiquetas CETEMPúblico / see CETEMPúblico tags)</p> <p><b>c) [emp="..."]</b> (apenas pt → en / pt → en only)</p> <p><b>N</b> - nominal / noun  <b>ADJ</b> - adjetival / adjective  <b>ADV</b> - adverbial / adverb  <b>PRP</b> - prepositiva / preposition  <b>KC</b> - coordenativa / coordinating  <b>KS</b> - subordinativa / subordinate</p>	<p>matches any common or proper noun  <b>NN</b> - nome singular / noun, singular or mass  <b>NNS</b> - nome plural / noun, plural  <b>NP</b> - nome próprio singular / proper noun, singular  <b>NPS</b> - nome próprio plural / proper noun, plural  <b>POS</b> - terminação de possessivo / possessive ending  <b>PP</b> - pronome pessoal / personal pronoun  <b>PP.*</b> - qualquer pronome pessoal ou possessivo / matches any pronoun - personal or possessive  <b>PP\$</b> - pronome possessivo / possessive pronoun  <b>RB</b> - advérbio / adverb  <b>RBR</b> - advérbio comparativo / adverb, comparative  <b>RBS</b> - advérbio superlativo / adverb, superlative  <b>SENT</b> - pontuação de fim de frase / sentence-final punctuation  <b>SYM</b> - símbolo / symbol  <b>TO</b> - qualquer uso de <i>to</i> / any use of 'to'  <b>UH</b> - interjeição / interjection  <b>VB</b> - verbo, forma base / verb, base form  <b>VB.*</b> - qualquer forma verbal / matches any verbal form  <b>VBD</b> - passado / verb, past tense  <b>VBG</b> - gerúndio / verb, gerund or present participle  <b>VBN</b> - particípio passado / verb, past participle  <b>WDT</b> - determinante iniciado por <i>wh</i> / wh-determiner  <b>WP</b> - pronome iniciado por <i>wh</i> / wh-pronoun  <b>WP\$</b> - possessivo / possessive  <b>WRB</b> - advérbio iniciado por <i>wh</i> / wh-adverb</p>
--	---	--

**Exercícios / Exercises**

**OPUS – EMEA (European Medicines Agency documents)**

- 1) What prepositions follow “impact”?
- 2) What prepositions follow the verb “fight”?
- 3) Search for “fast” as a noun, verb, adverb, but not as an adjective
- 4) Search for instances of up to 5 words separating “whether” and “or not”
- 5) Search for instances of up to 3 words between “approach” and a singular or plural noun followed by an infinitive with “to”

**COMPARA**

- 1) Search for a noun, followed by either “is” or “was”, followed by a verb ending in “-ed”