

Workshop: Como pesquisar em corpora

Sintaxe de pesquisa (Query syntax) CWB

O sistema IMS-CWB permite dois tipos de consulta:

1. **Concordâncias:** pedido de frases no corpus que contemplem a palavra/expressão de pesquisa.
2. **Distribuição:** pedido de quantificação/número de vezes que a palavra/expressão ocorre.

Alguns operadores de pesquisa:

Op.	Ex.	Descrição:
"..."	"casa"	As aspas balizam palavras e definem valores dos atributos.
[]	[lema="amar"]	Os parênteses rectos especificam os atributos das palavras/expressões.
()	"(a de)"	Agrupa lemas, palavras, expressões ou construções.
	"(a de)"	Operador de alternativa, disjunção.
!	[word!="..."]	Operador de negação; excluir o assinalado.
.	"pens."	Representa qualquer carácter (um só); exemplo: a palavra pode terminar em qualquer carácter: <i>penso</i> ; <i>pensa</i> .
*	"penso.*"	Quantificador de caracteres: zero ou mais vezes; ex: a palavra pode terminar por vários caracteres: <i>penso</i> , <i>pensamento</i> , <i>pensando</i> , etc.
?	"pobre(za)?"	Quantificador de caracteres: zero ou uma vez.
&	[word="casa" & pos="N"]	Operador de conjunção: junta ou combina atributos para especificar ou definir elementos de pesquisa (no exemplo, a palavra <i>casa</i> está definida enquanto 'nome').
=	[pos="N"]	Define o valor do atributo.

Projeto AC/DC	COMPARA	OPUS
<p>O sistema CWB foi utilizado no projeto AC/DC. Este projeto integra vários corpora e o utilizador pode pesquisar num corpus em particular ou proceder a uma procura simultânea em todos os corpora.</p> <p>Sintaxe de pesquisa utilizada no projeto AC/DC: http://www.linguateca.pt/ACDC/</p> <p>Atributos:</p> <p>word: palavra (forma) → [word="..."] lema: forma base da palavra que corresponde à 'unidade lexicográfica' → [lema="..."] pos (<i>part-of-speech tags</i>) categoria gramatical: nome, verbo, adjetivo, etc. temcagr: informação morfológica: tempo e modo verbal e caso para os pronomes → [temcagr="..."] pessnum: informação morfológica: pessoa e número → [pessnum="..."] gen: informação morfológica: género → [gen="..."] func: informação sintática → [func="..."] deriv: palavras derivadas → [deriv="DER.*"]</p> <p>Alguns valores:</p> <p>[pos="..."] ↓</p> <p>ADJ = adjetivo (ex: amável);</p> <p>ADV = advérbio;</p> <p>DET = determinante (artigo, pronome ou adjetivo); DEArtD = artigo definido (a, as, o, os); DEArtI = artigo indefinido (um, uns, uma, umas); DEAdem = demonstrativo; DEAposs = possessivo</p> <p>KS = conjunção subordinativa (ex: Não havia</p>	<p>A sintaxe de anotação morfossintática do COMPARA segue a utilizada no projeto AC/DC, no entanto foram detalhadas outras informações mais específicas; as várias categorias morfossintáticas estão etiquetadas com um conjunto de atributos e valores distintos do sistema de codificação de corpora IMS-CWB.</p> <p>Interface de pesquisa: http://www.linguateca.pt/COMPARA/</p> <p>Atributos:</p> <p>word → [word="..."] lema → [lema="..."] pos → [pos="..."] temcagr → [temcagr="..."] pessnum → [pessnum="..."] gen → [gen="..."] emp: locuções → [emp="..."]</p> <p>Alguns valores:</p> <p>[pos="..."] ↓</p> <p>ADJ = adjetivo (ex: <i>um homem de negócios suíço</i>); ADJn = adjetivo que funciona como nome (ex: <i>Nunca gostei muito da família real, mas é impossível não ter pena da <u>coitada</u> da rainha.</i>);</p> <p>ADV = advérbio;</p> <p>DET = determinante e pronome (menos os pessoais); DEArtD = artigo definido (a, as, o, os); DEArtI = artigo indefinido (um, uns, uma, umas); DEAdem = demonstrativo (ex: <i>Fiquei aliviado ao ouvir <u>estas</u> palavras.</i>);</p> <p>KC = conjunção coordenativa (<i>e, ou e mas</i>); KS = conjunção subordinativa ex: <i>Pensava que tu defendias <u>que</u> as mulheres devem ter o direito de</i></p>	<p>Interface de pesquisa: http://opus.lingfil.uu.se/bin/opuscqp.pl</p> <p>Atributos:</p> <p>word → [word="..."] lem (lema) → [lem="..."] pos → [pos="..."]</p> <p>Atributos de pos (inglês):</p> <p>[pos="..."] ↓</p> <p>CC = conjunção de coordenação (<i>coordinating conjunction</i>) CD = numeral cardinal (<i>cardinal number</i>) DT = determinante (<i>determiner</i>) EX = verbo <i>haver</i> (<i>existencial there</i>) FW = palavra estrangeira (<i>foreign word</i>) IN = preposição ou conjunção subordinativa (<i>preposition or subordinating conjunction</i>) JJ = adjetivo (adjective) JJR = adjetivo comparativo (<i>adjective, comparative</i>) JJS = adjetivo superlativo (<i>adjective, superlative</i>) MD = modo (<i>modal</i>) N.* = qualquer nome comum ou próprio (<i>matches any common or proper noun</i>) NN = nome singular (<i>noun, singular or mass</i>) NNS = nome plural (<i>noun, plural</i>) NP = nome próprio singular (<i>proper noun, singular</i>) NPS = nome próprio plural (<i>proper noun, plural</i>) POS = terminação de possessivo (<i>possessive ending</i>) PP = pronome pessoal (<i>personal pronoun</i>) PP.* = qualquer pronome pessoal ou possessivo (<i>matches any pronoun - personal or possessive</i>) PP\$ = pronome possessivo (<i>possessive pronoun</i>)</p>

<p><u>porque</u> repetir sem sentimento.); KC = conjunção coordenativa (mas; ou; e; etc);</p> <p>N = nome comum (ex: a <u>mulher</u>); PROP = nome próprio (ex: Lisboa);</p> <p>NUM = numeral (ex: As mulheres de <u>dezoito</u> a <u>trinta</u> anos são-lhe mais susceptíveis);</p> <p>PERS = pronome pessoal; PERS_refl = reflexo (me, te, se, si, nos, vos);</p> <p>PRP = preposição (ex: um homem <u>de</u> corpo forte e robusto);</p> <p>SPEC = especificador – pronomes (ex: <u>Ninguém</u> morreu de fome, nem de fastio.); SPEC_dem = demonstrativo (ex: <u>Isso</u> afinal é o menos.); SPEC_rel = relativo (ex: Falam-lhe no pai <u>que</u> morreu no patíbulo); SPEC_interr = interrogativo (ex: <u>Quem</u> procura o senhor?);</p> <p>V = verbo;</p> <p><u>Contrações:</u></p> <p>PERS\+PERS = pronomes pessoais (ex: As pessoas já <u>mo</u> têm dito.);</p> <p>PRP\+DET = preposição e determinante (ex: Os shows começam a partir <u>das</u> 16h30.);</p> <p>V\+PERS = verbo e pronome pessoal (ex: O senhor viu por acaso <u>tirarem-me</u> o chapéu?); V_fmcl\+PERS\+PERS = verbo de oração principal e pronomes (ex: Quando apareceu uma boneca de farrapos no leilão, ele <u>comprou-ma</u>.);</p> <p>temcagr: tempo e modo para os verbos, casos para os pronomes (PERS = pronomes pessoais e graus para os adjetivos</p> <p>[temcagr="..."] ↓</p> <p>PR_IND = Presente do Indicativo IMPF_IND = Imperfeito do Indicativo PS_IND = Pretérito Perfeito do Indicativo MQP_IND = Mais-Que-Perfeito do Indicativo FUT_IND = Futuro do Indicativo PR_SUBJ = Presente do Conjuntivo IMPF_SUBJ = Imperfeito do Conjuntivo FUT_SUBJ = Futuro do Conjuntivo IMP = Presente do Imperativo COND = Condicional INF = Infinitivo GER = Gerúndio PCP = Particípio</p> <p>Casos:</p> <p>NOM = nominativo (ex: <u>ela</u> sabia uma vez mais); ACC = acusativo (ex: a primeira vez que <u>o</u> viu); DAT = dativo (ex: como <u>lhe</u> competia);</p> <p>Graus:</p> <p>DIM = diminutivo – para nomes, adjetivos (ex: na água <u>azulinha</u>) AUM = aumentativo – para nomes e adjetivos (ex: com forro <u>azulão</u>)</p> <p>pessnum: pessoa/número = → [pessnum="..."] ↓</p> <p>número: S (singular) P (plural)</p>	<p>escolher.);</p> <p>N = nome comum (ex: Disse que queria o <u>carro</u>.); Nprop = nome comum iniciado por maiúscula (ex: recebendo o troféu das mãos da <u>Rainha</u>); PROP = nome próprio (ex: Teria preferido ir de carro a <u>Londres</u>); N_PROP = nome comum que pode ser também nome próprio (ex: Inverno)</p> <p>NUM = numeral (se não mesmo pelos <u>cinquenta</u> e poucos.); NUMcard = numeral cardinal (ex: Eram <u>9.30</u> da manhã.);</p> <p>PERS = pronome pessoal (ex: <u>ele</u> disse); PERSrefl = reflexo (me, te, se, si, nos, vos);</p> <p><u>Contrações com pronomes pessoais:</u></p> <p>PERS\+PERS = pronome pessoal e pronome pessoal (ex: Ela prometeu que <u>mo</u> pregava.); PERSrefl\+PERS = pronome pessoal reflexo e pronome pessoal (ex: como <u>no-lo</u> confirma o primeiro livro das Pandectas);</p> <p>PRP = preposição (ex: parecia uma pata <u>de</u> elefante);</p> <p><u>Contrações com preposições:</u></p> <p>PRP\+PERS = preposição e pronome pessoal (ex: Os outros tipos também repararam <u>nela</u>.); PRP\+PERSrefl = preposição e pronome pessoal reflexo (ex: Já venho ter <u>consigo</u>.);</p> <p>SPEC = pronome que funciona como nome (menos os pessoais); (ex: Não há <u>ninguém</u> que possa fazê-la tão bem como eu.); SPECdem = pronome demonstrativo (ex: Achei que já sabia <u>aquilo</u>.); SPECrel = pronome relativo (ex: há homens <u>que</u> tem os braços mais curtos);</p> <p>V = verbo;</p> <p><u>Contrações de verbos:</u></p> <p>V\+PERS = verbo e pronome pessoal (ex: Não pude <u>evitá-lo</u>.); V\+PERSrefl = verbo e pronome reflexo (ex: E sorriu, como se estivesse a <u>atribuir-me</u> um galardão.); V\+PERS\+PERS = verbo e pronomes pessoais contraídos (ex: A Amy não ficou nada impressionada quando tentei <u>explicar-lho</u>.);</p> <p>[temcagr="..."] ↓ (ver etiquetas projeto AC/DC)</p> <p>[emp="..."] (apenas pt → en) ↓</p> <p>N = nominal (ex: o <u>aquecimento central</u>) ADJ = adjetival (ex: um selo <u>de primeira classe</u>) ADV = adverbial (ex: não vai <u>a pé</u>) PRP = prepositiva (ex: <u>a partir de</u> terça-feira) KC = coordenativa (ex: <u>senão</u> que) KS = subordinativa (ex: a menos que)</p>	<p>RB = advérbio (<i>adverb</i>) RBR = advérbio comparativo (<i>adverb, comparative</i>) RBS = advérbio superlativo (<i>adverb, superlative</i>) SENT = pontuação de fim de frase (<i>sentence-final punctuation</i>) SYM = símbolo (<i>symbol</i>) TO = qualquer uso de to (<i>any use of 'to'</i>) UH = interjeição (<i>interjection</i>) VB = verbo – forma base (<i>verb, base form</i>) VB.* = qualquer forma verbal (<i>matches any verbal form</i>) VBD = passado (<i>verb, past tense</i>) VBG = gerúndio (<i>verb, gerund or present participle</i>) VCN = particípio passado (<i>verb, past participle</i>) WDT = determinante iniciado por wh (<i>wh-determiner</i>) WP = pronome iniciado por wh (<i>wh-pronoun</i>) WP\$ = possessive (<i>possessive</i>) WRB = advérbio iniciado por wh (<i>wh-adverb</i>)</p>
--	---	---

<p style="text-align: center;">SIP (singular e plural)</p> <p>peessoa: 1S ; 2S ; 3S (1ª, 2ª e 3ª pessoa do singular) 1P ; 2P ; 3P (1ª, 2ª e 3ª pessoa do plural) (1 3S = 1ª e 3ª pessoa do singular) (etc)</p> <p>gen:género = → [gen="..."] ↓ M (masculino) F (feminino) M F (indeterminado)</p> <p>deriv: palavras derivadas por prefixação ou sufixação → [deriv="DER.*"]</p> <p>Exemplo: pesquisar nomes (ex: <i>sacrossantidade</i>, <i>alminha</i>, <i>infelizardos</i>, <i>cartistas</i>) [pos="N.*" & deriv="DER.*"]</p>		
--	--	--

Exercícios

Projeto AC/DC: Corpus CETEMPúblico

1. Pesquise expressões com o verbo *dar*, no infinitivo, seguido de nome comum no singular seguido de preposição.
2. Pesquise ocorrências do verbo *estar* flexionado seguido da preposição *em* e de nome comum.

COMPARA

1. Pesquise equivalentes em português de advérbios terminados em “-wise”.
2. Pesquise equivalentes em inglês de locuções nominais que integrem a palavra *sala*.

Corpus OPUS

1. No subcorpus EMEA pesquise os equivalentes portugueses da estrutura inglesa composta por um advérbio iniciado por “wh-” seguido de um verbo no gerúndio.
2. No subcorpus EUROPARL 3 pesquise os equivalentes portugueses da expressão inglesa “swear an oath”, usando um comando que especifique as formas do lema “swear”.