# DCR-CORE Processing Results

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 10 | **document** | **document** | | container for document | dict | TET |
| 11 | document | createdAt | | created at (DD.MM.YY HH:MM:SS) | str | system |
| 11 | document | createdBy | | created by module | str | system |
| 11 | document | modifiedAt | x | last modified at (DD.MM.YY HH:MM:SS) | str | system |
| 11 | document | modifiedBy | x | last modified by module | str | system |
| 11 | document | noFonts | x | number fonts | int | TET |
| 11 | document | noFooters | | number footers per page | int | LT |
| 11 | document | noHeaders | | number headers per page | int | LT |
| 11 | document | noLines | | number lines in document | int | TET |
| 11 | document | noLinesFooter | | number footer lines | int | LT |
| 11 | document | noLinesHeader | | number header lines | int | LT |
| 11 | document | noLinesHeading | x | number heading lines | int | LT |
| 11 | document | noLinesListBullet | x | number bulleted list lines | int | LT |
| 11 | document | noLinesListNumber | x | number numbered list lines | int | LT |
| 11 | document | noLinesTable | | number table lines | int | LT |
| 11 | document | noLinesToc | x | number toc lines | int | LT |
| 11 | document | noListsBullet | x | number bulleted lists | int | LT |
| 11 | document | noListsNumber | x | number numbered lists | int | LT |
| 11 | document | noPages | | number pages in document | int | TET |
| 11 | document | noParas | | number paragraphs in document | int | TET |
| 11 | document | noSentences | x | number sentences in document | int | spaCy |
| 11 | document | noTables | | number tables in document | int | LT |
| 11 | document | noWords | | number words in document | int | TET |
| 19 | document | config | x | container for configuration | dict | TET |
| 19 | document | fonts | x | container for fonts | list | TET |
| 19 | document | headings | | container for headings | list | TET |
| 19 | document | listsBullet | | container for bulleted lists | list | TET |
| 19 | document | listsNumber | | container for numeredd lists | list | TET |
| 19 | document | pages | | container for pages | list | TET |
| 19 | document | params | x | container for parameters | dict | TET |
| 20 | **config** | **config** | x | container for configuration | dict | TET |
| 21 | config | parser | | container for module parser | dict | TT |
| 21 | config | tokenizer | | container for module tokenizer | dict | TET |
| 22 | config | jsonInclConfig | | include the configuration data in the JSON file | bool | setup |
| 22 | config | jsonInclFonts | | include the font data in the JSON file | bool | setup |
| 22 | config | jsonInclHeading | | include the heading data in the JSON file | bool | setup |
| 22 | config | jsonInclListBullet | | include the bulleted list data in the JSON file | bool | setup |
| 22 | config | jsonInclListNumber | | include the numbered list data in the JSON file | bool | setup |

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 22 | config | jsonInclParams | | include the parameters in the JSON file | bool | setup |
| 22 | config | jsonInclSentences | | include the sentences in the JSON file | bool | setup |
| 22 | config | jsonIndent | | Improves the readability of the JSON file | int | setup |
| 22 | config | jsonSortKeys | | Sort the keys in ascending order | bool | setup |
| 22 | config | ltFooterMaxDistance | | maximum Levenshtein distance for a footer line | int | setup |
| 22 | config | ltFooterMaxLines | | maximum number of footers | int | setup |
| 22 | config | ltHeaderMaxDistance | | Maximum Levenshtein distance for a header line | int | setup |
| 22 | config | ltHeaderMaxLines | | Maximum number of headers | int | setup |
| 22 | config | ltHeadingFileInclNoCtx | | number of lines following the heading to be included as context into the JSON file | bool | setup |
| 22 | config | ltHeadingFileInclRegexp | | number of lines following the heading to be included as context into the JSON file | bool | setup |
| 22 | config | ltHeadingMaxLevel | | maximum level of the heading structure | int | setup |
| 22 | config | ltHeadingMinPages | | minimum number of pages to determine the headings | int | setup |
| 22 | config | ltHeadingRuleFile | | file with rules to determine the headings | bool | setup |
| 22 | config | ltHeadingToleranceLlx | | tolerance of vertical indentation in percent | int | setup |
| 22 | config | ltListBulletMinEntries | | minimum number of entries to determine a bulleted list | int | setup |
| 22 | config | ltListBulletRuleFile | | file with rules to determine the bulleted lists | bool | setup |
| 22 | config | ltListBulletToleranceLlx | | tolerance of vertical indentation in percent | bool | setup |
| 22 | config | ltListNumberFileInclRegexp | | if it is set to true, the regular expression for the numbered list is included in the JSON file | bool | setup |
| 22 | config | ltListNumberMinEntries | | minimum number of entries to determine a numbered list | bool | setup |
| 22 | config | ltListNumberRuleFile | | file with rules to determine the numbered lists | bool | setup |
| 22 | config | ltListNumberToleranceLlx | | tolerance of vertical indentation in percent | bool | setup |

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 22 | config | ltTableFileInclEmptyColumns | | if it is set to true, the empty ells are included in the separate JSON file with the tables | bool | setup |
| 22 | config | ltTocLastPage | | maximum number of pages for the search of the TOC (from the beginning) | bool | setup |
| 22 | config | ltTocMinEntries | | minimum number of TOC entries | bool | setup |
| 22 | config | spacyIgnoreBracket | | ignore the tokens which are brackets ? | bool | setup |
| 22 | config | spacyIgnoreLeftPunct | | ignore the tokens which are left punctuation marks, e.g. "(" ? | bool | setup |
| 22 | config | spacyIgnoreLineTypeFooter | | ignore the tokens from line type footer ? | bool | setup |
| 22 | config | spacyIgnoreLineTypeHeader | | ignore the tokens from line type header ? | bool | setup |
| 22 | config | spacyIgnoreLineTypeHeading | | ignore the tokens from line type heading ? | bool | setup |
| 22 | config | spacyIgnoreLineTypeListBullet | | ignore the tokens from line type bulleted list ? | bool | setup |
| 22 | config | spacyIgnoreLineTypeListNumber | | ignore the tokens from line type numbered list ? | bool | setup |
| 22 | config | spacyIgnoreLineTypeTable | | ignore the tokens from line type table ? | bool | setup |
| 22 | config | spacyIgnoreLineTypeToc | | ignore the tokens from line type TOC ? | bool | setup |
| 22 | config | spacyIgnorePunct | | ignore the tokens which are punctuations ? | bool | setup |
| 22 | config | spacyIgnoreQuote | | ignore the tokens which are quotation marks ? | bool | setup |
| 22 | config | spacyIgnoreRightPunct | | ignore the tokens which are right punctuation marks, e.g. ")" ? | bool | setup |
| 22 | config | spacyIgnoreSpace | | ignore the tokens which consist of whitespace characters ? | bool | setup |
| 22 | config | spacyIgnoreStop | | ignore the tokens which are part of a “stop list” ? | bool | setup |
| 22 | config | spacyTknAttrCluster | | brown cluster ID | bool | setup |
| 22 | config | spacyTknAttrDep_ | | syntactic dependency relation | bool | setup |
| 22 | config | spacyTknAttrDoc | | the parent document | bool | setup |
| 22 | config | spacyTknAttrEntIob_ | | IOB code of named entity tag | bool | setup |
| 22 | config | spacyTknAttrEntKbId_ | | knowledge base ID that refers to the named entity this token is a part of, if any | bool | setup |

# DCR-CORE Processing Results

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 22 | config | spacyTknAttrEntType_ | | named entity type | bool | setup |
| 22 | config | spacyTknAttrHead | | the syntactic parent, or “governor”, of this token | bool | setup |
| 22 | config | spacyTknAttrI | | the index of the token within the parent document | bool | setup |
| 22 | config | spacyTknAttrIdx | | the character offset of the token within the parent document | bool | setup |
| 22 | config | spacyTknAttrIsAlpha | | does the token consist of alphabetic characters? | bool | setup |
| 22 | config | spacyTknAttrIsAscii | | does the token consist of ASCii characters? | bool | setup |
| 22 | config | spacyTknAttrIsBracket | | is the token a bracket? | bool | setup |
| 22 | config | spacyTknAttrIsCurrency | | is the token a currency symbol? | bool | setup |
| 22 | config | spacyTknAttrIsDigit | | does the token consist of digits? | bool | setup |
| 22 | config | spacyTknAttrIsLeftPunct | | is the token a left punctuation mark, e.g. "(" ? | bool | setup |
| 22 | config | spacyTknAttrIsLower | | is the token in lowercase? | bool | setup |
| 22 | config | spacyTknAttrIsOov | | is the token out-of-vocabulary? | bool | setup |
| 22 | config | spacyTknAttrIsPunct | | is the token punctuation? | bool | setup |
| 22 | config | spacyTknAttrIsQuote | | is the token a quotation mark? | bool | setup |
| 22 | config | spacyTknAttrIsRightPunct | | is the token a right punctuation mark, e.g. ")" ? | bool | setup |
| 22 | config | spacyTknAttrIsSentEnd | | does the token end a sentence? | bool | setup |
| 22 | config | spacyTknAttrIsSentStart | | does the token start a sentence? | bool | setup |
| 22 | config | spacyTknAttrIsSpace | | does the token consist of whitespace characters? | bool | setup |
| 22 | config | spacyTknAttrIsStop | | is the token part of a “stop list”? | bool | setup |
| 22 | config | spacyTknAttrIsTitle | | is the token in titlecase? | bool | setup |
| 22 | config | spacyTknAttrIsUpper | | is the token in uppercase? | bool | setup |
| 22 | config | spacyTknAttrLang_ | | language of the parent document’s vocabulary | bool | setup |
| 22 | config | spacyTknAttrLeftEdge | | the leftmost token of this token’s syntactic descendants | bool | setup |
| 22 | config | spacyTknAttrLemma_ | | base form of the token, with no inflectional suffixes | bool | setup |
| 22 | config | spacyTknAttrLex | | the underlying lexeme | bool | setup |

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 22 | config | spacyTknAttrLexId | | sequential ID of the token's lexical type, used to index into tables, e.g. for word vectors | bool | setup |
| 22 | config | spacyTknAttrLikeEmail | | does the token resemble an email address? | bool | setup |
| 22 | config | spacyTknAttrLikeNum | | does the token represent a number? | bool | setup |
| 22 | config | spacyTknAttrLikeUrl | | does the token resemble a URL? | bool | setup |
| 22 | config | spacyTknAttrLower_ | | lowercase form of the token text | bool | setup |
| 22 | config | spacyTknAttrMorph | | morphological analysis | bool | setup |
| 22 | config | spacyTknAttrNorm_ | | the token's norm, i.e. a normalized form of the token text | bool | setup |
| 22 | config | spacyTknAttrOrth_ | | verbatim text content | bool | setup |
| 22 | config | spacyTknAttrPos_ | | coarse-grained part-of-speech from the Universal POS tag set | bool | setup |
| 22 | config | spacyTknAttrPrefix_ | | a length-N substring from the start of the token | bool | setup |
| 22 | config | spacyTknAttrProb | | smoothed log probability estimate of token's word type | bool | setup |
| 22 | config | spacyTknAttrRank | | sequential ID of the token's lexical type, used to index into tables, e.g. for word vectors | bool | setup |
| 22 | config | spacyTknAttrRightEdge | | the rightmost token of this token's syntactic descendants | bool | setup |
| 22 | config | spacyTknAttrSent | | the sentence span that this token is a part of | bool | setup |
| 22 | config | spacyTknAttrSentiment | | a scalar value indicating the positivity or negativity of the token | bool | setup |
| 22 | config | spacyTknAttrShape_ | | transform of the token's string to show orthographic features | bool | setup |
| 22 | config | spacyTknAttrSuffix_ | | length-N substring from the end of the token | bool | setup |
| 22 | config | spacyTknAttrTag_ | | fine-grained part-of-speech | bool | setup |
| 22 | config | spacyTknAttrTensor | | the token's slice of the parent doc's tensor | bool | setup |
| 22 | config | spacyTknAttrText | | verbatim text content | bool | setup |
| 22 | config | spacyTknAttrTextWithWs | | text content, with trailing space character if present | bool | setup |

# DCR-CORE Processing Results

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 22 | config | spacyTknAttrVocab | | the vocab object of the parent doc | bool | setup |
| 22 | config | spacyTknAttrWhitespace_ | | trailing space character if present | bool | setup |
| 30 | **font** | **font** | x | container for font | dict | TET |
| 31 | font | embedded | | embedded | bool | TET |
| 31 | font | fontNo | | font number in document | int | TET |
| 31 | font | fullName | | font full name | str | TET |
| 31 | font | id | | font identification | str | TET |
| 31 | font | italicAngle | | font italic angle | float | TET |
| 31 | font | name | | font name | str | TET |
| 31 | font | type | | font type | str | TET |
| 31 | font | weight | | font weight | float | TET |
| 40 | **heading** | **heading** | x | container for heading entry | dict | LT |
| 41 | heading | ctxLine1 | | context line 1 | str | LT |
| 41 | heading | ctxLine2 | | context line 2 | str | LT |
| 41 | heading | ctxLine3 | | context line 3 | str | LT |
| 41 | heading | level | | heading level | int | LT |
| 41 | heading | lineNoPage | | line number in page | int | LT |
| 41 | heading | lineNoPara | | line number in paragraph | int | LT |
| 41 | heading | pageNo | | page number in document | int | LT |
| 41 | heading | paraNoPage | | paragraph number in page | int | LT |
| 41 | heading | regexp | x | regular expression | str | LT |
| 41 | heading | text | | heading text | str | LT |
| 50 | **listBullet** | **bulleted list** | x | container for bulleted list | dict | LT |
| 51 | listBullet | format | | bullet format | str | LT |
| 51 | listBullet | listNo | | list number in document | int | LT |
| 51 | listBullet | noEntries | | number entries in list | int | LT |
| 51 | listBullet | pageNoFirst | | page number first entry | int | LT |
| 51 | listBullet | pageNoLast | | page number last entry | int | LT |
| 52 | listBullet | entries | | container for list entries | list | LT |
| 53 | entryBullet | entry | | container for bulleted list entry | dict | LT |
| 53 | entryBullet | entryNo | | entry number in list | int | LT |
| 53 | entryBullet | lineNoPageFirst | | first line in page | int | LT |
| 53 | entryBullet | lineNoPageLast | | last line in page | int | LT |
| 53 | entryBullet | pageNo | | page number in document | int | LT |
| 53 | entryBullet | paraNo | | paragraph number in page | int | LT |
| 53 | entryBullet | text | | list entry text | str | LT |
| 60 | **listNumber** | **numbered list** | x | container for numbered list | dict | LT |
| 61 | listNumber | format | | bullet format | str | LT |
| 61 | listNumber | listNo | | list number in document | int | LT |
| 61 | listNumber | noEntries | | number entries in list | int | LT |
| 61 | listNumber | pageNoFirst | | page number first entry | int | LT |
| 61 | listNumber | pageNoLast | | page number last entry | int | LT |
| 61 | listNumber | regexp | x | regular expression | str | LT |
| 62 | listNumber | entries | | container for list entries | list | LT |
| 63 | entryNumber | entry | | container for numbered list entry | dict | LT |

# DCR-CORE Processing Results

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 63 | entryNumber | entryNo | | entry number in list | int | LT |
| 63 | entryNumber | lineNoPageFirst | | first line in page | int | LT |
| 63 | entryNumber | lineNoPageLast | | last line in page | int | LT |
| 63 | entryNumber | pageNo | | page number in document | int | LT |
| 63 | entryNumber | paraNo | | paragraph number in page | int | LT |
| 63 | entryNumber | text | | list entry text | str | LT |
| 70 | **page** | **page** | | container for page | dict | TET |
| 71 | page | lineNoFirst | | number of first line in page | int | TET |
| 71 | page | lineNoLast | | number of last line in page | int | TET |
| 71 | page | noLinesPage | | number of lines in page | int | TET |
| 71 | page | noParasPage | | number of paragraphs in page | int | TET |
| 71 | page | noWordsPage | | number of words in page | int | TET |
| 71 | page | pageNo | | page number in document | int | TET |
| 71 | page | paraNoFirst | | number of first paragraph in page | int | TET |
| 71 | page | paraNoLast | | number of last paragraph in page | int | TET |
| 71 | page | sentenceNoFirst | | number of first sentence in page | int | spaCy |
| 71 | page | sentenceNoLast | | number of last sentence in page | int | spaCy |
| 71 | page | wordNoFirst | | number of first word in page | int | TET |
| 71 | page | wordNoLast | | number of last word in page | int | TET |
| 79 | page | lines | | container for lines | list | TET |
| 79 | page | paras | | container for paragraphs | list | TET |
| 80 | **params** | **params** | x | container for parameters | dict | TET |
| 81 | params | parser | | container for module parser | dict | TT |
| 82 | params | directoryName | | file directory for intermediate files and final result files | str | params |
| 82 | params | documentId | x | document identification | int | params |
| 82 | params | environmentVariant | | environment variant | str | params |
| 82 | params | fileNameCurr | | name of input file | str | params |
| 82 | params | fileNameNext | | name of output file | str | params |
| 82 | params | fileNameOrig | | original file name | str | params |
| 82 | params | ltHeadingRequired | | heading determination required | bool | params |
| 82 | params | ltListBulletRequired | | bulleted list determination required | bool | params |
| 82 | params | ltListNumberRequired | | numbered list determination required | bool | params |
| 82 | params | ltTableRequired | | table determination required | bool | params |
| 82 | params | ltTocRequired | | TOC determination required | bool | params |
| 83 | params | tokenizer | | container for module tokenizer | dict | TET |
| 100 | **line** | **line** | | container for lines | dict | TET |
| 101 | line | level | x | heading level | int | LT |

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 101 | line | lineNo | | line number in document | int | TET |
| 101 | line | lineNoPage | | line number in page | int | TET |
| 101 | line | lineNoPara | | line number in paragraph | int | TET |
| 101 | line | llx | | x coordinate of the lower left corner | float | TET |
| 101 | line | noWordsLine | | number of words in line | int | TET |
| 101 | line | pageNo | | page number in document | int | TET |
| 101 | line | paraNo | | paragraph number in document | int | TET |
| 101 | line | paraNoPage | | paragraph number in page | int | TET |
| 101 | line | sentenceNo | | sentence number in document | int | spaCy |
| 101 | line | tableCellNo | | cell number in row | int | TET |
| 101 | line | tableCellSpan | | cel span | int | TET |
| 101 | line | tableNo | | table number in document | int | TET |
| 101 | line | tableRowNo | | row number in table | int | TET |
| 101 | line | text | | line text | str | TET |
| 101 | line | type | | line type | str | TET, TL |
| 101 | line | urx | | x coordinate of the upper right corner | float | TET |
| 101 | line | wordNoFirst | | number word of first word in line | int | TET |
| 101 | line | wordNoLast | | number word of last word in line | int | TET |
| 101 | line | wordNoParaFirst | | word number in paragraph of first word in this line | int | TET |
| **110** | **para** | **para** | | container for paragraph | dict | TET |
| 111 | para | lineNoFirst | | number of first line in paragraph | int | TET |
| 111 | para | lineNoLast | | number of last line in paragraph | int | TET |
| 111 | para | noLinesPara | | number of lines in paragraph | int | TET |
| 111 | para | noWordsPara | | number of words in paragraph | int | TET |
| 111 | para | pageNo | | page number in document | int | TET |
| 111 | para | paraNo | | paragraph number in document | int | TET |
| 111 | para | paraNoPage | | paragraph number in page | int | TET |
| 111 | para | sentenceNoFirst | | number of first sentence in paragraph | int | spaCy |
| 111 | para | sentenceNoLast | | number of last sentence in paragraph | int | spaCy |
| 111 | para | tableCellNo | | cell number in row | int | TET |
| 111 | para | tableCellSpan | | cel span | int | TET |
| 111 | para | tableNo | | table number in document | int | TET |
| 111 | para | tableRowNo | | row number in table | int | TET |
| 111 | para | text | | paragraph text | str | TET, spaCy |

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 111 | para | wordNoFirst | | number of first word in paragraph | int | TET |
| 111 | para | wordNoLast | | number of last word in paragraph | int | TET |
| 119 | para | sentences | x | container for sentences | list | spaCy |
| 119 | para | words | | container for words | list | TET |
| 120 | **sentence** | **sentence** | | container for sentences | dict | spaCy |
| 121 | sentence | sentenceNo | | sentence number in document | int | spaCy |
| 121 | sentence | sentenceNoPage | | sentence number in page | int | spaCy |
| 121 | sentence | sentenceNoPara | | sentence number in paragraph | int | spaCy |
| 121 | sentence | text | | sentence text | str | spaCy |
| 121 | sentence | wordNoFirst | | number word of first word in sentence | int | spaCy |
| 121 | sentence | wordNoLast | | number word of last word in sentence | int | spaCy |
| 200 | **word** | **word** | | container for word | dict | TET |
| 201 | word | font | x | font identification | str | TET |
| 201 | word | level | x | heading level | int | LT |
| 201 | word | lineNo | | line number in document | int | TET |
| 201 | word | lineNoPage | | line number in page | int | TET |
| 201 | word | llx | | x coordinate of the lower left corner | float | TET |
| 201 | word | pageNo | | page number in document | int | TET |
| 201 | word | paraNo | | paragraph number in document | int | TET |
| 201 | word | sentenceNo | | sentence  number in document | int | spaCy |
| 201 | word | size | x | font size | float | TET |
| 201 | word | spacyTknAttrCluster | | brown cluster ID | str | spaCy |
| 201 | word | spacyTknAttrDep_ | | syntactic dependency relation | str | spaCy |
| 201 | word | spacyTknAttrDoc | | the parent document | str | spaCy |
| 201 | word | spacyTknAttrEntIob_ | | IOB code of named entity tag | str | spaCy |
| 201 | word | spacyTknAttrEntKbId_ | | knowledge base ID that refers to the named entity this token is a part of, if any | str | spaCy |
| 201 | word | spacyTknAttrEntType_ | | named entity type | str | spaCy |
| 201 | word | spacyTknAttrHead | | the syntactic parent, or “governor”, of this token | str | spaCy |
| 201 | word | spacyTknAttrI | | the index of the token within the parent document | str | spaCy |
| 201 | word | spacyTknAttrIdx | | the character offset of the token within the parent document | str | spaCy |
| 201 | word | spacyTknAttrIsAlpha | | does the token consist of alphabetic characters? | bool | spaCy |

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 201 | word | spacyTknAttrIsAscii | | does the token consist of ASCii characters? | bool | spaCy |
| 201 | word | spacyTknAttrIsBracket | | is the token a bracket? | bool | spaCy |
| 201 | word | spacyTknAttrIsCurrency | | is the token a currency symbol? | bool | spaCy |
| 201 | word | spacyTknAttrIsDigit | | does the token consist of digits? | bool | spaCy |
| 201 | word | spacyTknAttrIsLeftPunct | | is the token a left punctuation mark, e.g. "(" ? | bool | spaCy |
| 201 | word | spacyTknAttrIsLower | | is the token in lowercase? | bool | spaCy |
| 201 | word | spacyTknAttrIsOov | | is the token out-of-vocabulary? | bool | spaCy |
| 201 | word | spacyTknAttrIsPunct | | is the token punctuation? | bool | spaCy |
| 201 | word | spacyTknAttrIsQuote | | is the token a quotation mark? | bool | spaCy |
| 201 | word | spacyTknAttrIsRightPunct | | is the token a right punctuation mark, e.g. ")" ? | bool | spaCy |
| 201 | word | spacyTknAttrIsSentEnd | | does the token end a sentence? | bool | spaCy |
| 201 | word | spacyTknAttrIsSentStart | | does the token start a sentence? | bool | spaCy |
| 201 | word | spacyTknAttrIsSpace | | does the token consist of whitespace characters? | bool | spaCy |
| 201 | word | spacyTknAttrIsStop | | is the token part of a “stop list”? | bool | spaCy |
| 201 | word | spacyTknAttrIsTitle | | is the token in titlecase? | bool | spaCy |
| 201 | word | spacyTknAttrIsUpper | | is the token in uppercase? | bool | spaCy |
| 201 | word | spacyTknAttrLang_ | | language of the parent document’s vocabulary | str | spaCy |
| 201 | word | spacyTknAttrLeftEdge | | the leftmost token of this token’s syntactic descendants | str | spaCy |
| 201 | word | spacyTknAttrLemma_ | | base form of the token, with no inflectional suffixes | str | spaCy |
| 201 | word | spacyTknAttrLex | | the underlying lexeme | str | spaCy |
| 201 | word | spacyTknAttrLexId | | sequential ID of the token’s lexical type, used to index into tables, e.g. for word vectors | str | spaCy |
| 201 | word | spacyTknAttrLikeEmail | | does the token resemble an email address? | bool | spaCy |
| 201 | word | spacyTknAttrLikeNum | | does the token represent a number? | bool | spaCy |
| 201 | word | spacyTknAttrLikeUrl | | does the token resemble a URL? | bool | spaCy |
| 201 | word | spacyTknAttrLower_ | | lowercase form of the token text | str | spaCy |
| 201 | word | spacyTknAttrMorph | | morphological analysis | str | spaCy |

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 201 | word | spacyTknAttrNorm_ | | the token's norm, i.e. a normalized form of the token text | str | spaCy |
| 201 | word | spacyTknAttrOrth_ | | verbatim text content | str | spaCy |
| 201 | word | spacyTknAttrPos_ | | coarse-grained part-of-speech from the Universal POS tag set | str | spaCy |
| 201 | word | spacyTknAttrPrefix_ | | a length-N substring from the start of the token | str | spaCy |
| 201 | word | spacyTknAttrProb | | smoothed log probability estimate of token's word type | str | spaCy |
| 201 | word | spacyTknAttrRank | | sequential ID of the token's lexical type, used to index into tables, e.g. for word vectors | str | spaCy |
| 201 | word | spacyTknAttrRightEdge | | the rightmost token of this token's syntactic descendants | str | spaCy |
| 201 | word | spacyTknAttrSent | | the sentence span that this token is a part of | str | spaCy |
| 201 | word | spacyTknAttrSentiment | | a scalar value indicating the positivity or negativity of the token | str | spaCy |
| 201 | word | spacyTknAttrShape_ | | transform of the token's string to show orthographic features | str | spaCy |
| 201 | word | spacyTknAttrSuffix_ | | length-N substring from the end of the token | str | spaCy |
| 201 | word | spacyTknAttrTag_ | | fine-grained part-of-speech | str | spaCy |
| 201 | word | spacyTknAttrTensor | | the token's slice of the parent doc's tensor | str | spaCy |
| 201 | word | spacyTknAttrText | | verbatim text content | str | spaCy |
| 201 | word | spacyTknAttrTextWithWs | | text content, with trailing space character if present | str | spaCy |
| 201 | word | spacyTknAttrVocab | | the vocab object of the parent doc | str | spaCy |
| 201 | word | spacyTknAttrWhitespace_ | | trailing space character if present | str | spaCy |
| 201 | word | tableCellNo | | cell number in row | int | TET |
| 201 | word | tableCellSpan | | cel span | int | TET |
| 201 | word | tableNo | | table number in document | int | TET |
| 201 | word | tableRowNo | | row number in table | int | TET |
| 201 | word | text | | word text | str | TET |
| 201 | word | type | | line type | str | TET, TL |
| 201 | word | urx | | x coordinate of the upper right corner | float | TET |
| 201 | word | wordNo | | word number in document | int | TET |
| 201 | word | wordNoLine | | word number in line | int | TET |

# DCR-CORE Processing Results

| lvl | container | tag | opt | description | type | source |
|---|---|---|---|---|---|---|
| 201 | word | wordNoPage | | word number in page | int | TET |
| 201 | word | wordNoPara | | word number in paragraph | int | TET |