**描述性统计【2项】**

|  |  |
| --- | --- |
| 单词数量 | DESWC（03） |
| 句子数量 | DESSC（02） |

Number of words (DESWC). (index 03)——文本单词总数

This is the total number of words in the text. Words are calculated using the output from the Charniak parser. For each sentence, the Charniak parser generates a parse tree with part of speech (POS) tags for clauses, phrases, words and punctuations. The elements on the leaves of a parse tree are tagged words or punctuations. In Coh-Metrix, words are taken from the leaves of the sentence parse trees.

Number of sentences (DESSC). (index 02)——文本句子总数

This is the total number of sentences in the text. Sentences are identified by the OpenNLP sentence splitter

**微观叙事结构分析【19项】**

维度一：词法（8项）

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 多样性 | LDVOCDa（51） |  |  |  |  |  |
| 密度 | WRDFRQC（94） |  |  |  |  |  |
| 复杂性 | K1 | K2 | K3 | K4 | K5 | K6+ |

LDVOCDa (index 51)——数值代表词汇多样性，数值越大越多样

VOC lexical diversity measure for all words.

WRDFRQc (index 94)——文本中“实义词”的平均词频

This is the average word frequency for content words.

K1/K2/K3/K4/K5/K6+

K级代表文本单词的复杂程度，K级别越高代表所用单词越高级。

维度二：句法（11项）

|  |  |  |  |
| --- | --- | --- | --- |
| 短语长度指标 | SYNLE（69） | SYNNP（70） |  |
| 短语密度指标 | DRNP（76） | DRVP（77） | DRPP（79） |
| 句子长度指标 | DESSL（06） |  |  |
| 句子结构指标 | DRPVAL（80） | DRNEG（81） | DRINF（83） |
| 句法相似指标 | SYNSTRUTa（74） | SYNSTRUTt（75） |  |

Words before main verb: SYNLE (index 69)——句子中，主要动词前的平均单词数量

This is the mean number of words before the main verb of the main clause in sentences. This is a good index of working memory load.

Modifiers per NP: SYNNP (index 70)——每个名词短语中修饰语的平均数量

This is the mean number of modifiers per noun-phrase.

DRNP (index 76)——名词短语发生率得分

This is the incidence score of noun phrases.

DRVP (index 77)——动词短语发生率得分

This is the incidence score of verb phrases.

DRPP (index 79)——介词短语 发生率得分

This is the incidence score of preposition phrases.

Mean number of words (length) of sentences in (DESSL). (index 06)——句子的平均单词数量

This is the average number of words in each sentence within the text, where a word is anything that is tagged as a part-of-speech by the Charniak parser. Sentences with more words may have more complex syntax and may be more difficult to process. While this is a descriptive measure, this also provides one commonly used proxy for syntactic complexity. However, Coh-Metrix provides additional more precise measures of syntactic complexity discussed later in this chapter.

DRPVAL (index 80)——无主语被动语态发生率得分

This is the incidence score of agentless passive voice forms.

Negations: DRNEG (index 81)——否定式表达发生率得分

This is the incidence score for negation expressions.

DRINF (index 83)——不定式发生率得分

This is the incidence score of infinitives.

Syntactic structure similarity adjacent: SYNSTRUTa (index 74)

——所有相邻句子的句法结构相似性

This is the proportion of intersection tree nodes between all adjacent sentences.

Syntactic structure similarity all 01: SYNSTRUTt (index 75)

——所有句子和段落的句法结构相似性

This is the proportion of intersection tree nodes between all sentences and across paragraphs.

**宏观叙事结构分析【13项】**

维度一：语篇连贯性（8项）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 局部连贯性 | CRFAO1（29） | CRFSO1（30） | LSASS1（40） |  |  |
| 整体连贯性 | CRFAOa（32） | CRFSOa（33） | LSASSp（42） | LSAPP1（44） | LSAGN（46） |

Argument overlap (CRFAO1 and CRFAOa). (index 29, 32)

29——相邻论元重叠；32——论元重叠

These local and global overlap measures are similar to noun overlap measures, but include overlap between sentences in terms of nouns and pronouns. Argument overlap occurs when there is overlap between a noun in one sentence and the same noun (in singular or plural form) in another sentence; it also occurs when there are matching personal pronouns between two sentences (e.g., he/he). The term argument is used in a linguistic sense, where noun/pronoun arguments are contrasted with verb/adjective predicates (Kintsch & Van Dijk, 1978). Consider argument overlap for the science passage in Table 4.1 in the second column. Note that in comparison to noun overlap, it is less strict because it considers the overlap for example between cells and cell. Argument and stem overlap would also include overlap between pronouns, such as it to it, or he to he, which noun overlap does not include.

Stem overlap (CRFSO1, CRFSOa). (index 30, 33)

30——相邻词干重叠；33——词干重叠

These two local and global overlap measures relax the noun constraint held by the noun and argument overlap measures. A noun in one sentence is matched with a content word (i.e., nouns, verbs, adjectives, adverbs) in a previous sentence that shares a common lemma (e.g., tree/treed; mouse/mousey; price/priced). Notably, the outcome for stem and argument overlap in Table 4.1 were identical; however, this will not always be the case.

LSA sentence adjacent: LSASS1(index 40)——相邻句子间 潜在语义分析

This index computes mean LSA cosines for adjacent, sentence-to-sentence (abbreviated as "ass") units. This measures how conceptually similar each sentence is to the next sentence.

LSA sentence all: LSASSp (index 42)——所有句子间 潜在语义分析

Like LSA sentence adjacent (LSAassa), this index computes mean LSA cosines. However, for this index all sentence combinations are considered, not just adjacent sentences. LSApssa computes how conceptually similar each sentence is to every other sentence in the text.

LSAPP1 (index 44)——段落间 潜在语义分析

This index computes the mean of the LSA cosines between adjacent paragraphs.

LSAGN (index 46)——新旧信息 潜在语义分析

This is the avarage givenness of each sentence.

维度二：情境模型（3项）

|  |  |
| --- | --- |
| 因果衔接 | SMCAUSr（64） |
| 意图衔接 | SMINTEr（65） |
| 时体衔接 | SMTEMP（68） |

Causal cohesion: SMCAUSr (index 64)——因果衔接度

This is a ratio of causal particles (P) to causal verbs (V). The denominator is incremented by the value of 1 to handle the rare case when there are 0 causal verbs in the text. Cohesion suffers when the text has many causal verbs (signifying events and actions) but few causal particles that signal how the events and actions are connected.

Intentional cohesion: SMINTEr (index 65)——意图衔接度

This is the ratio of intentional particles to intentional actions/events.

Temporal cohesion: SMTEMP (index 68)——时体衔接度

This is the repetition score for tense and aspect. The repetition score for tense is averaged with the repetition score for aspect.

维度三：段落长度（2项）**\*后续需要配合SG Model质性分析**

|  |  |
| --- | --- |
| 段落平均长度 | DESPL（04） |
| 段落平均长度标准差 | DESPLd（05） |

Mean length of paragraphs (DESPL). (index 04)——段落平均长度

This is the average number of sentences in each paragraph within the text. Longer paragraphs may be more difficult to process.

Standard deviation of the mean length of paragraphs (DESPLd). (index 05)

——段落平均长度标准差

This is the standard deviation of the measure for the mean length of paragraphs within the text. In the output, d is used at the end of the name of the indices to designate that it is a standard deviation. A large standard deviation indicates that the text has large variation in terms of the lengths of its paragraphs, such that it may have some very short and some very long paragraphs. The presence of headers in a short text can increase values on this measure.