### **Assessing Text Difficulty**

### Department of Modern Languages Research and Development & Testing Units

Department of Modern Languages (MLD) Testing Committee uses two computer tools, Coh-Metrix version 3.0 and Lexical Tutor version 8 VocabProfile when assessing the difficulty of the reading texts in the midterm and final exams. In order to be able to make valid comparisons, consistency is essential when using these tools (Ürkün, 2014). In addition to the data provided by these tools, test writers evaluate the texts intuitively considering certain text characteristics that are not likely to be evaluated accurately by the available computer systems.

Currently, the difficulty level of the texts in the coursebook is used to set the baseline in the evaluation and selection of the texts to be used in the exams (See Appendix A for the Coh-Metrix version 3.0 Indices for the reading texts in the coursebook, *Compass 1*). It is also assumed that students are at B2 level and above since they passed ODTU English Proficiency Exam (EPE) or a similar proficiency exam recognized by ODTU.

#### 1. Using Coh-Metrix version 3.0

Coh-Metrix version 3.0 is a computational tool which "analyses texts on multiple measures of language and discourse that are aligned with multilevel theoretical frameworks of comprehension" (Graesser, McNamara, & Kulikowich, 2011, p. 223). There are a number of reasons why Coh-Metrix version 3.0 is used for assessing text readability:

- It is "easy to use and free" (Elfenbein, 2011, p.246)
- It allows uploading texts up to 15,000 words in length (Dowell, Graesser, & Cai, 2015)
- Coh-Metrix is based on psycholinguistic and cognitive models of reading. (Crossley, Allen, & McNamara, 2011), and compared to traditional readability formulas such as Flesch-Kincaid, which scales texts on "a single metric of text ease or difficulty", it is a more valid and reliable tool in assessing readability (Graesser et al., 2011, p. 224). Coh-Metrix version 3.0 provides data on 108 indices.

On the other hand, there are some difficulties of using Coh-Metrix. First, as stated above, the program provides information on 108 indices, which makes it a challenge to compare texts. As Graesser & Elfenbein state "texts equated on number of syllables, word frequency, type-token ratio and syntactic difficulty still may vary on other measures, including causal relatedness, semantic overlap, and emotional content" (Elfenbein, 2011, p. 246). Second, as indicated on the webpage, "the scores are often subject to the output of third party parsers, lexicons, and word frequency databases, all of which are outside of the control of Coh-Metrix" (McNamara, Louwerse, Cai, & Graesser, 2013, Preliminary Information section, para. 1). Finally, there are some text features that might be beyond the scope of Coh-Metrix:

Available computer systems cannot fully comprehend the deep metaphors, literary devices, and historical contexts of Shakespeare's plays, for example. Some characteristics of texts require humans to provide informed, deep critical analyses. Second, successful text comprehension involves much more than an analysis of text characteristics alone because prior knowledge, inference mechanisms, and skills of readers are also critically important. (Graesser et al., 2011, p. 223)

To be able to use the program effectively, selecting indices that are more likely to predict text difficulty is important. One option is to use Coh-Metrix TEA tool (http://tea.cohmetrix.com), which is a compact version of Coh-Metrix; however, since this program works only with texts up to 1000 words in length, it is not suitable for use with the texts used in MLD exams (Dowell et al., 2015). MLD testing committee conducted a comprehensive literature review to determine the indices to focus on when assessing text difficulty. This report includes the findings of this review and the implications for MLD.

The designers of Coh-Metrix refer to 5 principal Coh-Metrix components; *narrativity*, *deep cohesion*, *referential cohesion*, *syntactic simplicity*, and *word concreteness*: (Dowell et al., 2015):

• **Narrativity.** Narrative text tells a story, with characters, events, places, and things that are familiar to the reader. Narrative is closely affiliated with everyday oral conversation. This robust component is highly affiliated with word familiarity, world knowledge, and oral language. Informational expository texts on less familiar topics would lie at the opposite end of the continuum. (Index PCNARp, *CohMx13*)

• **Deep cohesion**. This dimension reflects the degree to which the text contains causal, intentional, and temporal connectives and conceptual links. These connectives help the

reader to form a more coherent and deeper understanding of the causal events, processes, and actions in the text. (Index PCDCp, *CohMx21*)

• **Referential cohesion.** This component includes Coh-Metrix indices that assess referential cohesion. High cohesion text contains words and ideas that overlap across sentences and the entire text, forming explicit threads that connect the text for the reader. Low cohesion text is typically more difficult to process because there are fewer threads that tie the ideas together for the reader. (Index PCREFp, *CohMx19*)

•Syntactic Simplicity. This component reflects the degree to which the sentences in the text contain fewer words and use more simple, familiar syntactic structures, which are less challenging to process. At the opposite end of the continuum are texts that contain sentences with more words, embedded constituents, unfamiliar syntactic structures, noun-phases with many modifiers, and many words before the main verb of the main clause (i.e., left-embedded syntax that is taxing on working memory). (Index PCSYNp, *CohMx15*)

• Word Concreteness. Texts that contain content words that are concrete, meaningful, and evoke mental images are easier to process and understand. Abstract words represent concepts that are difficult to represent visually. Texts that contain more abstract words are more challenging to understand. (Index PCCNCp, *CohMx17*)

These principal indices are taken into consideration in readability assessment. However, it is important to note that cohesion indices should be used with caution (Ürkün, 2014). In her workshop, Ürkün cited a number of research studies that support this point:

•recent literature suggests that different forms of cohesion are not always positively correlated with grade-level bands (Graesser et al., 2011)

•in studies carried out by CRELLA, cohesion indices did not clearly relate to different levels of text either in reading or writing.

Green, Khalifa and & Weir (2013) conducted a study to identify which Coh-Metrix variables were better predictors of text difficulty. Using Coh-Metrix 2.0 (at the time they carried out the study, there were 54 indices on Coh-Metrix) and VocabProfile, they analysed 116 reading texts used in *Cambridge English: First* (B2), *Cambridge English: Advanced* (C1) and *Cambridge English: Proficiency* (C2) exams to determine text characteristics changing by

level. Statistical analysis revealed 15 indices on Coh-Metrix, and Academic Wordlist (AWL) and Offlist words on VocabProfile as differentiating characteristics. These 15 indices are listed and explained below. Both labels in the version the researchers used in their study and their equivalents in Coh-Metrix Version 3.0 are given. One index that could not be matched in the two versions is indicated with a minus (-). The indices are explained briefly using the information provided on the webpage of Coh-Metrix and in literature.

Green et al. (2013) identified three lexical variables that distinguished the texts used at three different levels of the Cambridge exams. Table 1 shows these three differentiating lexical indices.

Table 1

## Differentiating lexical indices

Coh-Metrix 2.0	Coh-Metrix 3.0				
CohMx38 Average syllables per word	Mean number of syllables (length) in words				
	(DESWLsy). (index 08)				
	Word length, number of syllables, mean				
	*CohMx8				
CohMx42 Higher level constituents per word	-				
CohMx44 Type-token ratio for all content words	Type-token ratio: LDTTRc (index 48)				
	Lexical diversity, type-token ratio, content word				
	lemmas CohMx46				
CohMx46, CELEX, logarithm, mean for content	WRDFRQc (index 94)				
words	CELEX word frequency for content words, mean				
	CohMx92				

\* The labels printed in red refer to the numbers provided at the output page of Coh-Metrix analysis

## Lexical indices explained:

## Mean number of syllables (length) in words (DESWLsy): (index 08)

"Coh-Metrix calculates the average number of syllables in all of the words in the text. Shorter words are easier to read and the estimate of word length serves as a common proxy for word frequency" (McNamara et al., 2013, Descriptive Indices section).

**Type-token ratio: (LDTTRc): (index 48)** McNamara et al. (2013, Lexical Diversity section) explain type-token ratio as follows:

Type-token ratio (TTR) (Templin, 1957) is the number of unique words (called types) divided by the number of tokens of these words. Each unique word in a text is considered a word type. Each instance of a particular word is a token. For example, if the word *dog* appears in the text 7 times, its type value is 1, whereas its token value is 7. When the type-token ratio approaches 1, each word occurs only once in the text; comprehension should be comparatively difficult because many unique words need to be decoded and integrated with the discourse context. As the type-token ratio decreases, words are repeated many times in the text, which should increase the ease and speed of text processing. Type-token ratios are computed for content words, but not function words. TTR scores are most valuable when texts of similar lengths are compared.

"When lexical diversity is at a maximum, the text is either very low in cohesion or perhaps the text is very short" (Dowell et al., 2015, para. 1)

### **CELEX** word frequency for content words, mean (WRDFRQc): (index 94)

"This is the average word frequency for content words" (McNamara et al., 2013, Word Information section).Texts containing high proportion of low-frequency words will be more difficult to process than those containing only very common words" (Green et al, 2013, p. 31). Therefore, the higher the score, the easier is processing (N. Dowell, personal communication, November 9, 2015).

As shown in Table 2, Green et al. (2013) found three differentiating syntactic indices.

## Table 2

Differentiating syntactic indices

Coh-Metrix 2.0	Coh-Metrix 3.0			
CohMx27 LSA, sentence to sentence	LSA sentence adjacent: LSASS1(index 40)			
adjacent mean	LSA overlap, adjacent sentences, mean			
	*CohMx38			
CohMx37 Average words per sentence	Mean number of words (length) of			
	sentences in (DESSL). (index 06)			
	Sentence length, number of words, mean			
	СоһМхб			
CohMx56 Sentence syntax similarity, all	Syntactic structure similarity all 01:			
across paragraphs	SYNSTRUTt (index 75)			
	Sentence syntax similarity, all combinations,			
	across paragraphs, mean CohMx73			

\* The labels printed in red refer to the numbers provided at the output page of Coh-Metrix analysis

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.

Example:

Text 1: The field was full of lush, green grass. The horses grazed peacefully. The young children played with kites. The women occasionally looked up, but only occasionally. A warm summer breeze blew and everyone, for once, was almost happy.

Text 2: The field was full of lush, green grass. An elephant is a large animal. No-one appreciates being lied to. What are we going to have for dinner tonight?

In the example texts printed above, Text 1 records much higher LSA scores than Text 2. The words in Text 1 tend to be thematically related to a pleasant day in an idyllic park scene: green, grass, children, playing, summer, breeze, kites, and happy, In contrast, the sentences in Text 2 tend to be unrelated.

Higher LSA scores lead to easier text processing.

### Mean number of words (length) of sentences in (DESSL). (index 06)

This index is explained as below on the Coh-Metrix webpage:

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. (McNamara et al., 2013, Descriptive Indices section)

Sentences with more words are more difficult to process than sentences with fewer words.

### Syntactic structure similarity all 01: SYNSTRUTt (index 75)

Green et al. (2013) explains this index as follows:

[The index in Coh-Metrix compares] "the syntactic tree structure of sentences. An issue what is known as a syntactic priming effect. It is well attested in language production research (Pickering and Branigan 1999) that after a speaker has formulated a particular syntactic structure, there is likelihood that they will employ a similar structure in the following utterance. This phenomenon is less clearly attested in reading comprehension. While syntactic priming appears to play a positive role in comprehension, it has been suggested that the effect may be partly or wholly due to the repetition of the verb. However, recent neurological evidence (Ledoux, Traxler and Saab 2007) suggest that syntactic parsing effects may be present even when the verb is not repeated. (p. 32)

However, Dowel warns the users of the program when making readability assumptions using this index. She states that "since this is a measure of similarity, it does not necessarily indicate simplicity. For example, if all sentences are syntactically complex, they would have a high similarity score, but [the text will] be harder to process" (N. Dowell, personal communication, November 9, 2015).

The study revealed 6 differentiating text-level representation indices. These indices are listed on Table 3.

# Table 3

# Differentiating text-level representation indices

Coh-Metrix 2.0	Coh-Metrix 3.0				
CohMx16 argument overlap, adjacent,	Argument overlap (CRFAO1). (index 29)				
unweighted	Argument overlap, adjacent sentences,				
	binary, mean *CohMx29				
CohMx18 Anaphor reference, adjacent,	Anaphor overlap (CRFANP1, CRFANPa)				
unweighted	(index 38,39) <i>CohMx?</i>				
CohMx21 Anaphor reference, all distances					
CohMx26 Logical operator incidence score	CNCLogic (index 54)				
	Logical connectives incidence CohMx 52				
CohMx58 Proportion of content words that	Content word overlap (CRFCWO1d).				
overlap between adjacent sentences	(Index 35)				
	Content word overlap, adjacent sentences,				
	proportional, mean CohMx34				
CohMx60 Concreteness, minimum in	WRDFRQmc (index 96)				
sentence for content words	CELEX Log minimum frequency for content				
	words, mean CohMx 94				

\* The labels printed in red refer to the numbers provided at the output page of Coh-Metrix analysis

# Text-level representation indices explained:

# Argument overlap (CRFAO1). (index 29)

"This index is the proportion of all sentence pairs per paragraph that share one or more arguments (i.e noun, pronoun, noun-phrase). A higher score is indicative of a more cohesive text and easier reading" (Green et al, 2013, p. 32).

## Anaphor overlap (CRFANP1, CRFANPa) (index 38,39)

McNamara et al. (2013, Referential Cohesion section) provide the explanation below on the Coh-Metrix webpage:

This measure considers the anaphor overlap between pairs of sentences. A pair of sentences has an anaphor overlap if the later sentence contains a pronoun that refers to

a pronoun or noun in the earlier sentence. The score for each pair of sentences is binary, i.e., 0 or 1. The measure of the text is the average of the pair scores. This measure includes both local (CRFANP1) and global (CRFANPa) indices.

Since these identities are not reported on the current publically available analysis of Coh-Metrix 0.3, they will not be included in readability assessment.

### **CNCLogic** (index 54)

Logical connectives is a subcategory of the connectives index. Connectives are significant "in establishing situation model cohesion (or Deep Cohesion)" (Dowell et al., 2015, Coh-Metrix Principal Components section):

Coh-Metrix delivers a relative frequency (index) score (occurrence per 1000 words) for all connectives as well as different types of connectives. Indices are provided on five broad categories of connectives: causal (because, so), additive (and, moreover), temporal (first, until), logical (and, or), and adversative/contrastive (although, whereas) which Coh--Metrix classifies on the basis of prior research (Halliday & Hasan, 1976; Louwerse, 2001). Additionally, Coh-Metrix differentiates between positive connectives (also, moreover) and negative connectives (however, but).

"CNCLogic is the incidence score of logic connectives" (McNamara et al., 2013, Connectives section). Green et al. (2013) state that logic connective "include *and*, *or*, *not*, *if*, *then* and a small number of other similar cognate terms" (p. 33).

### Content word overlap (CRFCWO1d). (Index 35)

"The Coh-Metrix index content word overlap, which measures how often content words overlap between two adjacent sentences, measures one of many factors that facilitate meaning construction" (Crossley, Greenfield & McNamara, 2008, p. 483). McNamara (2013, Referential Cohesion section) explains this index as follows:

This measure considers the proportion of explicit content words that overlap between pairs of sentences. For example, if a sentence pair has fewer words and two words overlap, the proportion is greater than if a pair has many words and two words overlap. This measure includes both local (CRFCWO1) and global (CRFCWOa) indices, and also includes their standard deviations (CRFCWO1d, CRFCWOad)...

This measure may be particularly useful when the lengths of the sentences in the text are a principal concern.

"Overlapping vocabulary has been found to be an important aspect in reading processing and can lead to gains in text comprehension and reading speed" (Crossley et al., 2008, p. 483). As Green et al. (2013) indicate, "the occurrence of the same content word in adjacent sentences reduces text difficulty" (p. 33).

#### WRDFRQmc (index 96)

"This is the average minimum word frequency in sentences" (McNamara et al., 2013, Word Information section). This Coh-Metrix measure is based upon the well-established finding that abstract words are more difficult to process because they are not as imaginable as concrete words" (Green et al., 2013, p. 33).

Green et al. (2013) further ran a multiple regression analysis to identify which of the sets and individual features were good predictors of text difficulty, and concluded that "features of cohesion (logical operator incidence, lexical overlap between sentences) and lexis (word frequency and the occurrence of infrequent and academic words) rather than syntax are criterial in distinguishing between the texts used in the three highest levels of the Cambridge English Examinations.

Another research study was done by Crossley, Greenfield and McNamara (2008). They investigated which Coh-Metrix indices were better predictors of text difficulty, with the hypothesis that "variables relating to lexical frequency, syntactic similarity, and content word overlap" would have a significant effect on text readability (Crossley et al, 2008, p. 481). In their research, they used 31 of the academic texts in Bormuth's corpus and the mean scores of the cloze tests prepared using these texts from an earlier study done with Japanese L2 learners. Based on L1 and L2 literature, they identified *CELEX frequency score for written words* index (Index 94, WRDFRQc CohMx 93. (Note that based on authors' other related work, Index 94 rather than Index 95, WRDFRQa, CohMx 93 is selected), *semantic similarity: sentence to sentence, adjacent, mean* index (Index 72, SYNSTRUTa, CohMx 72) and *content word overlap* index (CRFCWO1d, Index 35, *CohMx34*) as indices to be examined. *CELEX word frequency* and *content word overlap* indices are already explained below. Below Crossley et al.'s explanation of the *sentence semantic similarity* index is provided:

Semantic similarity: sentence to sentence, adjacent, mean measures explained: [This index is the] uniformity and consistency of parallel syntactic constructions in only looks at syntactic similarity at the phrase level, but also text. The index not takes account of the parts of speech involved, on the assumption that the more uniform the syntactic constructions are, the easier the syntax will be to process. It is important to include a measure of difficulty that is not simply based on the traditional L2 grading of grammar pat- terns but also takes account of how the reader handles words as they are encountered on the page. A reading text is processed linearly, with the reader decoding it word by word; but, as he or she reads, the reader also has to assemble decoded items into a larger scale syntactic structure (Just & Carpenter, 1987; Rayner & Pollatsek, 1994). Clearly, the cognitive demands imposed by this operation vary considerably according to how complex the structure is (Perfetti, Landi, & Oakhill, 2005). They also vary according to how predictable the final part of the structure is because, while still in the course of reading a sentence, we form expectations as to how it will end. So-called garden path sentences such as John remembered the answer / was in the book impose particularly heavy demands and contribute significantly to text difficulty (Field, 2004, pp. 121, 299). These factors of potential difficulty are provided for by the Coh-Metrix sematic similarity index. (2008, pp. 482-483)

They indicated that "reiterated syntactic structures lower the cognitive demands placed on L2 learners and afford them the opportunity to concentrate on meaning construction" (Crossley et al., 2008, pp. 489).

The results of the multiple regression analysis showed that the combination of *content word overlap, syntactic similarity*, and *CELEX frequency* accounted "for 86% of the variance in the performance of the Japanese students on the 31 cloze tests based on the Bormuth passages. In other words, using these three variables, the model can predict 86% of the difficulty for these passages" (Crossley et al., 2008, pp. 485).

Crossley, Allan and McNamara (2011) compared Coh-Metrix Second Language L2 with traditional readability formulas in order to identify which one better distinguished beginner, intermediate and advanced texts, and found that Coh-Metrix Second Language L2 performed better than the traditional formulas. In their study, they used a set of intuitively simplified texts and analysed them using both Coh-Metrix Second Language L2 and

traditional formulas, with the assumption that Coh-Metrix would better predict the intuitive classifications since intuitive approaches to text simplification are likely to take into account text characteristics considered important in psycholinguistic and cognitive reading models, which are overlooked in traditional readability formulas. Coh-Metrix L2 Reading Index is calculated using three indices in Coh-Metrix, CELEX Word Frequency (logarithm mean for content words), Sentence Syntax Similarity (sentence to sentence adjacent mean), and Content Word Overlap (proportional adjacent sentences unweighted) (Crossley et al., 2011). These indices are already explained earlier in the report. Statistical analysis revealed that Flesch Reading Ease classified 44% of the texts accurately. This percentage was 48% for the Flesch-Kincaid Grade Level formula, and Coh-Metrix classified 59% of the texts accurately. Although Coh-Metrix outperformed the traditional formula, it should be noted that only 59% of the classifications were accurate. The authors provided several explanations for this. First of all, Coh-Metrix L2 index includes only three of the variables on the Coh-Metrix tool. Second, the classifying the intermediate texts accurately using the formula were difficult. However, there is also the possibility that Coh-Metrix L2 Reading Index does not measure some features considered in the intuitive approach:

It is also likely that many of the intuitive simplification features in the texts that lead to better text comprehension were not measured by the Coh-Metrix L2 Reading Index. Such an assumption rests on the notion that the reading index only considers three variables, while the process of intuitive text simplification likely modifies a much larger number of linguistic features. Such an assumption does not challenge the strength of the Coh-Metrix L2 Reading Index, especially when compared to traditional readability formulas, but it does suggest that more research is needed to develop formulas that contain more linguistic features and that better match text readability for various genres, readers, and levels. (p. 98)

Higher Coh-Metrix L2 Reading Index scores indicate easier reading.

In the workshop, Ürkün (2015) shared some Coh-Metrix indices for Cambridge Main Suit Exams from a study conducted by Taylor and Weir in 2012. This data is given in Table 4 and 5. The comparison may be useful for test writers.

## Table 4

Coh-Metrix Index 6	Sentence	length,	number	of word	ls for	Cambridge	Texts
		0 /		9		0	

Main Suite Level	Average number of words	Range
	per sentence	
KET (A2)	13.2	8-17
PET (B1)	14.9	10-20
FCE (B2)	18.4	11-25
CAE (C1)	18.6	13-27
CPE (C2)	19.6	13-30

## Table 5

Coh-Metrix indices 104 & 105 Flesch Reading Ease Score & Flesch-Kincaid Grade Level for Cambridge texts

Main Suite Level	Flesch Reading	Flesch-Kincaid Grade	Flesch-Kincaid
	Ease Score	Level	Range
KET (A2)	78.3	5.5	2-7.4
PET (B1)	64.7	7.9	5-10.1
FCE (B2)	66.5	8.4	5-12.3
CAE (C1)	58.4	9.6	5.7-16
CPE (C2)	57.7	9.9	5.6-16.1

## **Guidelines for using Coh-Metrix**

## 1. Cleaning the data

In order to get valid analysis results using Coh-Metrix, the entered data should be clean. Dowell et al. (2015) state that "a clean text looks exactly like it would appear if the writer had just finished typing it, had it checked for typos and errors by a large group of copy editors, printed if off, and then handed it to the researcher" (The Corpus, Pre-Processing, and Best Practices for Text Analytic section). The data should be cleaned from annotations, odd

line breaks, spelling mistakes. They also note that data transferred from one computer to another can be polluted:

Similarly, corpora that have been passed around from computer to computer tend to "grow" various oddities such as the odd Spanish letter, or a string of mathematical symbols. Particularly in cases where researchers have converted documents that include pictures into text files, the pictures in the document disappear often leaving the caption of the pictures lurking oddly in the middle of the text. (Dowell et al., 2015, The Corpus, Pre-Processing, and Best Practices for Text Analytic section)

As they stress "each of these dirties has the potential to seriously undermine the validity of Coh-Metrix analyses" (Dowell et al., 2015, The Corpus, Pre-Processing, and Best Practices for Text Analytic section). They also make some suggestions on how to clean the data and offer two golden best practices to users:

- 1. If there is not a good reason to take it out, the researcher should leave it in.
- 2. What the researcher does to one text, should be done to all.

Best practice 1 basically states that the default condition of the text is exactly the way the researcher found it. Therefore, all changes made to it after that should be documented and reported for future replications. Most commonly, researchers decide to remove annotations and picture captions [emphasis added]. The logic behind this decision is that they make the text unreadable, and consequentially any Coh-Metrix results are likely to be seriously flawed. A different motivation might be reported for removing the picture captions. Here a strong argument would be that they are not part of the continuous text that the writer intended. Additionally, their insertion into the document renders the sentence meaningless, and the corresponding evaluations will be misleading. Best practice 2 is extremely important. It means a researcher should never pick and choose the texts that are modified. If something is removed from one text (e.g., a day, month, and year that happens to be at the end of a text) then one must confirm that none of the other texts also have that pattern (and if they do, they must all be removed, or all kept). Similarly, the same consistency should be used for spelling corrections and typos. Finally, it is important to understand that a few dirties across the corpus is not considered unusual. As a general rule of thumb, we say that the corpus needs to be at least 95% clean. That is, about 95% of the texts should have no problems at all, and at least 95% of each text should be thoroughly correct. When researchers have very large corpora, reading through all of them is not feasible. Note that in this context, assessing a random sample of the text (e.g. 10%) is generally considered sufficient.

Another way to clean the data is to use software. Some alternatives are the Coh-Metrix Data Viewer facility on Test Easability Assessor (TEA) (<u>http://tea.cohmetrix.com</u>), Notepad ++ (<u>http://notepad-plus-plus.org</u>) and Text Crawler (<u>http://textcrawler.en.softonic.com</u>). (Dowell et al., 2014). They also note that Coh-Metrix works best with Firefox.

## 2. Recording the data

Based on the literature review, a number of indices will be of primary focus when assessing text readability using Coh-Metrix. These indices are given in Table 6.

Table 6

## Coh-Metrix indices to be assessed in text selection

No	Label	Label V2.x	Full description							
Descrip	otive									
1	DESPC	READNP	Paragraph count, number of paragraphs							
2	DESSC	READNS	Sentence count, number of sentences							
3	DESWC	READNW	Word count, number of words							
4	DESPL	READAPL	Paragraph length, number of sentences in a paragraph, mean							
5	DESPLd	n/a	Paragraph length, number of sentences in a paragraph, standard deviation							
6	DESSL	READASL	Sentence length, number of words, mean							
7	DESSLd	n/a	Sentence length, number of words, standard deviation							
8	DESWLsy	READASW	Word length, number of syllables, mean							
9	DESWLsyd	n/a	Word length, number of syllables, standard deviation							
10	DESWLlt	n/a	Word length, number of letters, mean							
11	DESWLltd	n/a	Word length, number of letters, standard deviation							

Table 6 continued

No	Label	Label V2.x	Full description					
Text Easability Principle Component Scores								
No	Label	Label V2.x	Full description					
12	PCNARz	n/a	Text Easability PC Narrativity, z score					
13	PCNARp	n/a	Text Easability PC Narrativity, percentile					
14	PCSYNz	n/a	Text Easability PC Syntactic simplicity, z score					
15	PCSYNp	n/a	Text Easability PC Syntactic simplicity, percentile					
16	PCCNCz	n/a	Text Easability PC Word concreteness, z score					
17	PCCNCp	n/a	Text Easability PC Word concreteness, percentile					
18	PCREFz	n/a	Text Easability PC Referential cohesion, z score					
19	PCREFp	n/a	Text Easability PC Referential cohesion, percentile					
20	PCDCz	n/a	Text Easability PC Deep cohesion, z score					
21	PCDCp	n/a	Text Easability PC Deep cohesion, percentile					
Referential Cohesion								
29	CRFAO1	CRFBA1um	Argument overlap, adjacent sentences, binary, mean					
34	CRFCW01	CRFPC1um	Content word overlap, adjacent sentences,					
			proportional, mean					
35	CRFCW01d	n/a	Content word overlap, adjacent sentences,					
			proportional, standard deviation					
LSA								
38	LSASS1	LSAassa	LSA overlap, adjacent sentences, mean					
39	LSASS1d	LSAassd	LSA overlap, adjacent sentences, standard deviation					
Lexica	l Diversity							
46	LDTTRc	ТҮРТОКс	Lexical diversity, type-token ratio, content word					
			lemmas					
Connee	ctives							
52	CNCLogic	CONLOGi	Logical connectives incidence					
Syntactic Complexity								
72	SYNSTRUTa	STRUTa	Sentence syntax similarity, adjacent sentences, mean					
73	SYNSTRUTt	STRUTt	Sentence syntax similarity, all combinations, across					
			paragraphs, mean					

Table 6 continued

No	Label	Label V2.x	Full description						
Word I	Word Information								
92	WRDFRQc	FRCLacwm	CELEX word frequency for content words, mean						
94	WRDFRQmc	FRCLmcsm	CELEX Log minimum frequency for content words,						
			mean						
Readat	oility								
104	RDFRE	READFRE	Flesch Reading Ease						
105	RDFKGL	READFKGL	Flesch-Kincaid Grade level						
106	RDL2	L2	Coh-Metrix L2 Readability						

For the indices summarized in Table 6, the features of the texts used in the exams will be within the ranges of the features of the coursebook texts. Table 7 displays the mean scores and the ranges for Coh-Metrix indices for the texts in *Compass 1*. The results of the complete Coh-Metrix analysis of the texts in Compass 1 can be found in Appendix A.

Table 7

The mean scores and the ranges for Coh-Metrix indices for the texts in Compass 1

No	Description	Mean	Range
1	Paragraph Count	10.12	8-14
2	Number of Sentences	63.75	53-92
3	Word Count	1190.37	1000-1477
4	Paragraph length, number of sentences in a paragraph, mean	6.34	5.09-7.77
6	Sentence length, number of words, mean	18.94	16.05-21.23
8	Word length, number of syllables, mean	1.63	1.54-1.74
10	Word length, number of letters, mean	4.96	4.74-5.32
13	Text Easability PC Narrativity, percentile	33.00	17.11-52.39
15	Text Easability PC Syntactic simplicity, percentile	40.16	29.81-54.38
17	Text Easability PC Word concreteness, percentile	39.78	17.11-66.28
19	Text Easability PC Referential cohesion, percentile	10.35	5.48-26.43
21	Text Easability PC Deep cohesion, percentile	68.35	31.92-86.65

Table 7 continued

No	Description	Mean	Range
29	Argument overlap, adjacent sentences, binary, mean	0.41	0.31-0.53
34	Content word overlap, adjacent sentences,	0.06	0.04-0.08
	proportional mean		
38	LSA overlap, adjacent sentences, mean	0.15	0.09-0.19
46	Lexical diversity, type-token ratio, content word	0.67	0.63-0.7
	lemmas		
52	Logical connectives incidence	39.93	25.52-48.16
72	Sentence syntax similarity, adjacent sentences,	0.07	0.05-0.08
	mean		
73	Sentence syntax similarity, all combinations, across	0.06	0.04-0.07
	paragraphs, mean		
92	CELEX word frequency for content words, mean	2.19	2.14-2.25
94	CELEX Log minimum frequency for content words,	1.23	0.88-1.46
	mean		
104	Flesch Reading Ease	49.23	37.65-60.00
105	Flesch Kincaid Grade Level	11.09	8.87-13.28
106	Coh-Metrix L2 Readability	11.27	8.33-13.02

### 2. Using Lexical Tutor version 8 VocabProfile

MLD Testing Committee used Lexical Tutor version 8 VocabProfile in assessing text readability and text modification. Nakata (2013) explains how the software is used:

Vocabulary profilers refer to computer programs that compare a text against word lists specified by the user. These programs allow users to create their own frequency lists, identify words that are or are not shared by various texts, find words that are likely to be unknown to students of certain vocabulary knowledge, evaluate students' productive vocabulary knowledge, or estimate the vocabulary load of materials (Webb 2008). Compleat Lexical Tutor & Nation, also provides RANGE (www.lextutor.ca/range/) and Vocabulary Profilers (www.lextutor.ca/vp/). The difference between RANGE and Vocabulary Profilers is that the former can analyze multiple text files simultaneously while the latter can process only one at a time. (p. 8) In her study, Dodigovic (2005) explored the suitability of the texts used in two courses designed for non-native speakers of English at an international university, with the aim of explaining the low success and motivation levels of the learners in these courses. Using Lexical Tutor, she compared the students' vocabulary size with the vocabulary range in the textbooks. Based on Hirsh & Nation's (1992) and Laufer's (1989, 1992) work on vocabulary, she held "the assumption that the students should, at any given time, be able to understand 90–95% of the required readings in the program in order for comprehension to take place, which would then result in both content and vocabulary learning. (Dodigovic, 2005, p. 449). Thus, she focused on the three lists on Lexical Tutor; that is, the first 1000 (K1), the second 1000 (K2) and the AWL. She found that "15% of the vocabulary used in the texts was not found on any of the three lists" (p. 452). In her conclusion, Dodigovic highlighted the importance of considering the amount of unknown words in text selection:

When the required readings contain 5% more of the unusual vocabulary, the comprehension level is likely to go down to 75% or less of the text, thus making the message of the text unintelligible. Such a context would be expected to severely hamper the learning of both vocabulary and content. (Dodigovic, 2005, p. 452)

Dodigovic's study has important implications for MLD testing policy. Using Vocabulary Profile, the frequency of the words will be calculated and words above 4K will be screened and when necessary, they will be replaced or given in the gloss so that at least 95% of the words will be familiar to the test-takers. (Ürkun, 2015) In the future, it is suggested to create a corpora using the words in the texts in the course books.

#### **Guidelines for using VocabProfiler**

#### 1. Choosing the corpus

There are a number of corpora options on VocabProfiler. Among these, the Testing Committee uses BNC-COCA 1-25k. BNC is the British National Corpus and COCA is the Corpus of Contemporary American English. One of the reasons why this corpora is selected is that although still widely used, GSL corpora "was compiled more than half a century ago, it does not contain modern words such as television, computer, or online" (Nakata, 2013, p. 2). Unfortunately, this corpora does not provide a comparison with the AWL, which should be a point to consider when the curriculum incorporates an AWL teaching component.

2. Preparing the text and interpreting the results

Before entering the text, all the proper nouns in the text should be eliminated. It is possible to replace them with a word like PROPER. VocabProfiler (Cobb, 2015, Output page) explains how the data is processed by the software as stated below:

In the output text, punctuation is eliminated; all figures (1, 20, etc.) are replaced by the word number; contractions are replaced by constituent words (won't => will not); type-token ration is calculated using said constituents; and in the 1k sub-analysis content + function words may sum to less than total (depending on user treatment of proper nouns and program decision to class numbers as 1k although not contained in 1k list).

This is important to note when making comparisons among the results provided by different software programs.

One of the things that should be assessed is the type-token ratio. The closer this ratio is to 1, the more diverse is the lexical items in the text. Another point to consider is the percentage of the words at different K levels. Assuming that ENG 101 students are at B2 level in reading skills, 95% of the vocabulary should be at 4K (Ürkün, 2014). The software also provides a colour-coded analysis of words at each level and offlist words. It is possible to screen the words at each level and make modifications in the text. However, it should be noted that some of the offlist words can be familiar vocabulary. To exemplify how VocabProfiler is used, one sample analysis is provided below. The text is selected from *Compass 1*.

Sample Analysis:

Title: The Eyes Have it: Guess Who Controls the Future of TV?





Offlist words: au\_[1] classmates\_[1] contraire\_[1] dropbox\_[1] filmmakers\_[1] handheld\_[1] immersive\_[1] laptop\_[1] laptops\_[1] olds\_[1] ott\_[1] prosumer\_[2] sharable\_[1] smartphone\_[2] smartphones\_[1] widespread\_[1]

As can be seen, some of the offlist words such as *smartphone* (*s*), *laptop*(*s*), *dropbox*, *filmmakers*, *classmates*, *widespread* and *olds* are indeed words students familiar with, and OTT is an abbreviation already explained in the text.

### 3. Rerunning the modified text

If the text is simplified, it should be run through the program again in order to see how the modifications changed the vocabulary profile of the text.

During the research process, all the texts in *Compass 1* were analysed using VocabProfiler. However, the vocabulary in the book will not be used as the baseline when preparing the texts since the book itself does not specify a proficiency level for the users, and it is not based on a specific AWL. However, when the words in the offlist are target words specified in the coursebook, these words will not be simplified.

### 3. Intuitive Assessment

Although automated assessment of readability provides a systematic and scientific data, "pooled expert judgment is still necessary for some decisions, e.g. cultural specificity,

content knowledge, topic familiarity" (Ürkün, 2014). Alderson (2000) classifies variables that affect the nature of reading into two broad categories: reader variables and text variables. The reader variables "include the readers' background and subject/topic knowledge, their cultural knowledge, and their knowledge of the language in which the target text is written" (p. 80). When selecting texts for the exams, the first three variables are intuitively assessed by the test writers. These variables are important for a number of reasons:

1. Test takers will find it more difficult to read texts about unfamiliar topics.

2. Texts should not contain culturally biased and/ or offensive content.

3. If the text is too familiar in terms of topic, this also creates problems for assessment. Alderson explains how text familiarity can threaten the validity of an exam:

The importance of background, cultural, subject and topical knowledge in comprehension means that test designers must be aware that such knowledge may well influence test scores or measures of reading. Normally we are not interested in measuring such knowledge in reading tests: this would represent a reduction in the validity of our measure. One precaution, then, can be select texts on topics which are known to be equally familiar or unfamiliar to all candidates. The obvious problem is finding such topics. (p. 81)

Two other important reader variables are motivation and anxiety. In terms of motivation, Alderson suggests choosing texts that are likely to generate positive responses than negative responses. Choosing texts that are at the right linguistic level is also important in order not to demotivate students.

#### References

Alderson, J. C. (2000). Assessing reading. US: New York. Cambridge University Press.

- Cobb, T. *VocabProfile v.8* [computer program]. Accessed 15 Sept 2015 at http://www.lextutor.ca/cgi-bin/range/texts/index.pl
- Crossley, S. a., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42(3), 475 – 493. http://doi.org/10.1002/j.1545-7249.2008.tb001
- Crossley, S., Allen, D., & McNamara, D. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language*, *23*(1), 84–101.
- Dodigovic, M. (2005). Vocabulary profiling with electronic corpora: A case study in computer assisted needs analysis. *Computer Assisted Language Learning*, 18(5), 443– 455. http://doi.org/10.1080/09588220500442806
- Dowell, N., Cai, Z., & Graesser, A. C. (July, 2014). Analyzing language and discourse with Coh-Metrix. Workshop presented at 2nd Learning Analytics Summer Institutes (LASI 2014), Cambridge, MA
- Dowell, N. M. M., Graesser, A. C., & Cai, Z. (2015). Language and discourse analysis with Coh-Metrix : Applications from educational material to learning environments at scale. *Journal of Learning Analytics*, (October).
- Elfenbein, A. (2011). Research in text and the uses of Coh-Metrix. *Educational Researcher*, 40(5), 246–248. http://doi.org/10.3102/0013189X11414181
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234. http://doi.org/10.3102/0013189X11413260
- Green, A., Khalifia, H. & Weir, C. (2013). Examining textual features of reading texts. *Cambridge English Language Assessment Research Notes*, 52, 24-39
- Gülen, G., Hasanbaşoğlu, B., Şeşen, E. & Tokdemir, G. (2014). *The compass 1*. Türkiye, Ankara: NÜANS Publishing
- Nakata, T. (2013). Web-based lexical resources. *The Encyclopedia of Applied Linguistics*, 6166–6177. http://doi.org/10.1002/9781405198431.wbeal1275

- McNamara, D.S., Louwerse, M.M., Cai, Z., & Graesser, A. (2013). *Coh-Metrix version 3.0 indices*. Accessed 15 October 2015 at http://cohmetrix.com
- Ürkün, Z. (March, 2015). Assessment of receptive skills. 3 day training program conducted at the Department of Modern Languages, METU, Ankara.

## Appendix A: Table 8 Coh-Metrix indices for the reading texts in Compass 1

1. Your masterpiece yourself 2. Transhumanism 3. Online Identities 4. Social Media: The New Power of the public sphere 5. The Future of Reading in Online Revolution 6. The Eyes Have it: Guess Who Controls the Future of TV? 7. Kids Today 8. Boomerang Kids Rely on their Parents: Is It a Positive Trend?

No.	Label	Label V2.x	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Full description
	Descriptive										
1	DESPC	READNP	8	11	10	9	14	9	9	11	Paragraph count, number of paragraphs
2	DESSC	READNS	53	56	59	53	92	69	70	58	Sentence count, number of sentences
<mark>3</mark>	DESWC	READNW	1000	1189	1241	1079	1477	1142	1293	1102	Word count, number of words 1200- 1500
4	DESPL	READAPL	6.625	5.091	5.9	5.889	6.571	7.667	7.778	5.273	Paragraph length, number of sentences in a paragraph, mean
<mark>5</mark>	DESPLd	n/a	2.669	2.071	2.685	1.691	3.502	2.958	4.631	2.687	Paragraph length, number of sentences in a paragraph, sd
<mark>6</mark>	DESSL	READASL	18.868	21.232	21.034	20.358	16.054	16.551	18.471	19	Sentence length, number of words, mean

No.	Label	Label V2.x	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Full description
7	DESSLd	n/a	10.256	11.595	10.95	12.348	13.228	9.799	11.643	10.614	Sentence length, number of words, sd
<mark>8</mark>	DESWLsy	READASW	1.617	1.745	1.679	1.702	1.543	1.625	1.599	1.575	Word length, number of syllables, mean
<mark>9</mark>	DESWLsyd	n/a	0.948	1.04	1.028	1.054	0.853	0.951	0.922	0.862	Word length, number of syllables, sd
10	DESWLlt	n/a	4.864	5.326	5.021	5.134	4.741	4.9	4.797	4.95	Word length, number of letters, mean
<mark>11</mark>	DESWLltd	n/a	2.707	3.024	2.861	2.884	2.501	2.827	2.689	2.519	Word length, number of letters, sd
			Tex	t Easabil	ity Princi	ple Com	ponent So	cores		-	
12	PCNARz	n/a	-0.39	-0.768	-0.256	-0.953	-0.285	-0.603	-0.507	0.068	Text Easability PC Narrativity, z score
<mark>13</mark>	PCNARp	n/a	34.83	22.36	40.13	17.11	38.97	27.43	30.85	52.39	Text Easability PC Narrativity, percentile
14	PCSYNz	n/a	-0.537	0.115	-0.517	-0.365	-0.015	-0.015	-0.212	-0.532	Text Easability PC Syntactic simplicity, z score
<mark>15</mark>	PCSYNp	n/a	29.81	54.38	30.5	35.94	49.6	49.6	41.68	29.81	Text Easability PC Syntactic simplicity, percentile

No.	Label	Label V2.x	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Full description
16	PCCNCz	n/a	-0.333	-0.952	-0.239	-0.47	0.041	-0.596	-0.105	0.42	Text Easability PC Word concreteness, z score
<mark>17</mark>	PCCNCp	n/a	37.07	17.11	40.52	31.92	51.6	27.76	46.02	65.91	Text Easability PC Word concreteness, percentile
18	PCREFz	n/a	-1.444	-1.592	-0.631	-1.053	-1.6	-1.591	-1.333	-1.377	Text Easability PC Referential cohesion, z score
<mark>19</mark>	PCREFp	n/a	7.49	5.59	26.43	14.69	5.48	5.59	9.18	8.53	Text Easability PC Referential cohesion, percentile
20	PCDCz	n/a	0.379	0.929	0.291	0.667	0.295	1.11	-0.477	1.009	Text Easability PC Deep cohesion, z score
<mark>21</mark>	PCDCp	n/a	64.43	82.12	61.41	74.54	61.41	86.65	31.92	84.13	Text Easability PC Deep cohesion, percentile
22	PCVERBz	n/a	0.675	-0.349	-0.302	0.871	0.748	0.083	0.156	-0.225	Text Easability PC Verb cohesion, z score
23	PCVERBp	n/a	74.86	36.32	38.21	80.78	77.04	53.19	55.96	41.29	Text Easability PC Verb cohesion, percentile
24	PCCONNz	n/a	-2.835	-2.92	-1.837	-2.41	-1.733	-3.032	-1.894	-1.806	Text Easability PC Connectivity, z score

No.	Label	Label V2.x	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Full description
25	PCCONNp	n/a	0.23	0.18	3.36	0.8	4.18	0.12	2.94	3.59	Text Easability PC Connectivity, percentile
26	PCTEMPz	n/a	-0.099	-0.544	-0.392	-0.995	0.404	0.63	-0.551	0.736	Text Easability PC Temporality, z score
27	РСТЕМРр	n/a	46.41	29.46	34.83	16.11	65.54	73.57	29.12	76.73	Text Easability PC Temporality, percentile
			Ref	ferential (	Cohesion						
28	CRFNO1	CRFBN1um	0.308	0.236	0.397	0.442	0.231	0.191	0.275	0.246	Noun overlap, adjacent sentences, binary, mean
<mark>29</mark>	CRFAO1	CRFBA1um	0.423	0.418	0.5	0.538	0.319	0.338	0.377	0.421	Argument overlap, adjacent sentences, binary, mean
30	CRFSO1	CRFBS1um	0.404	0.509	0.483	0.558	0.319	0.279	0.348	0.298	Stem overlap, adjacent sentences, binary, mean
31	CRFNOa	CRFBNaum	0.156	0.123	0.312	0.236	0.131	0.181	0.177	0.175	Noun overlap, all sentences, binary, mean
32	CRFAOa	CRFBAaum	0.251	0.259	0.45	0.328	0.228	0.265	0.276	0.373	Argument overlap, all sentences, binary, mean

No.	Label	Label V2.x	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Full description
33	CRFSOa	CRFBSaum	0.211	0.329	0.428	0.368	0.254	0.258	0.24	0.234	Stem overlap, all sentences, binary, mean
<mark>34</mark>	CRFCW01	CRFPC1um	0.062	0.063	0.085	0.075	0.047	0.047	0.061	0.064	Content word overlap, adjacent sentences, proportional, mean
35	CRFCWO1d	n/a	0.082	0.086	0.099	0.085	0.086	0.082	0.097	0.091	Content word overlap, adjacent sentences, proportional, sd
36	CRFCWOa	CRFPCaum	0.038	0.03	0.059	0.045	0.031	0.041	0.04	0.052	Content word overlap, all sentences, proportional, mean
37	CRFCWOad	n/a	0.068	0.054	0.085	0.072	0.071	0.074	0.073	0.079	Content word overlap, all sentences, proportional, sd
			LS	A							
<mark>38</mark>	LSASS1	LSAassa	0.16	0.19	0.185	0.19	0.094	0.158	0.14	0.14	LSA overlap, adjacent sentences, mean
39	LSASS1d	LSAassd	0.154	0.136	0.148	0.144	0.131	0.105	0.138	0.105	LSA overlap, adjacent sentences, sd
40	LSASSp	LSApssa	0.136	0.172	0.183	0.162	0.085	0.162	0.114	0.127	LSA overlap, all sentences in paragraph, mean
41	LSASSpd	LSApssd	0.147	0.126	0.149	0.138	0.128	0.125	0.13	0.117	LSA overlap, all sentences in paragraph, sd

No.	Label	Label V2.x	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Full description
42	LSAPP1	LSAppa	0.334	0.312	0.433	0.466	0.321	0.522	0.294	0.332	LSA overlap, adjacent paragraphs, mean
43	LSAPP1d	LSAppd	0.139	0.129	0.127	0.139	0.147	0.128	0.105	0.153	LSA overlap, adjacent paragraphs, sd
44	LSAGN	LSAGN	0.291	0.29	0.32	0.293	0.279	0.301	0.283	0.286	LSA given/new, sentences, mean
45	LSAGNd	n/a	0.097	0.095	0.112	0.11	0.131	0.091	0.092	0.09	LSA given/new, sentences, sd
			Lex	ical Dive	rsity						
<mark>46</mark>	LDTTRe	TYPTOKe	0.705	0.68	0.639	0.678	0.679	0 645	0.69	0.665	Lexical diversity, type-token ratio,
		11110ite	0.705	0.00	0.057	0.070	0.077	0.015	0.09	0.005	content word lemmas
47	LDTTRa	n/a	0.458	0.448	0.393	0.441	0.423	0.426	0.438	0.436	Lexical diversity, type-token ratio, all words
48	LDMTLD	LEXDIVTD	119.617	123.13	94.469	94.669	101.084	113.919	101.585	143.86	Lexical diversity, MTLD, all words
49	LDVOCD	LEXDIVVD	123.665	145.057	122.694	105.628	119.608	112.785	114.932	132.196	Lexical diversity, VOCD, all words
			Con	inectives							
50	CNCAll	CONi	85	89.151	78.163	92.678	76.506	93.695	76.566	82.577	All connectives incidence
51	CNCCaus	CONCAUSi	24	26.913	22.562	29.657	20.311	30.648	15.468	30.853	Causal connectives incidence
<mark>52</mark>	CNCLogic	CONLOGi	40	46.257	44.319	38.925	37.238	48.161	25.522	39.02	Logical connectives incidence

No.	Label	Label V2.x	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Full description
53	CNCADC	CONADVCONi	27	22.708	17.728	14.829	12.187	25.394	13.921	18.149	Adversative and contrastive connectives incidence
54	CNCTemp	CONTEMPi	12	18.503	12.087	17.609	15.572	13.135	20.108	20.871	Temporal connectives incidence
55	CNCTempx	CONTEMPEXi	17	16.821	16.116	21.316	18.28	32.399	20.108	19.964	Expanded temporal connectives incidence
56	CNCAdd	CONADDi	49	50.463	42.707	55.607	41.3	53.415	44.084	36.298	Additive connectives incidence
57	CNCPos	n/a	0	0	0	0	0	0	0	0	Positive connectives incidence
58	CNCNeg	n/a	0	0	0	0	0	0	0	0	Negative connectives incidence
			Situ	ation Mo	odel						
59	SMCAUSv	CAUSV	17	23.549	19.339	23.17	20.988	20.14	25.522	20.871	Causal verb incidence
60	SMCAUSvp	CAUSVP	28	37.847	28.203	34.291	30.467	34.151	33.256	33.575	Causal verbs and causal particles incidence
61	SMINTEp	INTEi	10	7.569	14.504	10.195	12.864	9.632	19.335	16.334	Intentional verbs incidence
62	SMCAUSr	CAUSC	0.611	0.586	0.44	0.462	0.438	0.667	0.294	0.583	Ratio of casual particles to causal verbs
63	SMINTEr	INTEC	2	2.1	1.158	2.25	1.05	2.667	0.462	1.263	Ratio of intentional particles to intentional verbs

No.	Label	Label V2.x	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Full description
64	SMCAUSIsa	CAUSLSA	0.114	0.126	0.087	0.08	0.082	0.078	0.093	0.061	LSA verb overlap
65	SMCAUSwn	CAUSWN	0.626	0.383	0.518	0.644	0.575	0.482	0.476	0.534	WordNet verb overlap
66	SMTEMP	TEMPta	0.817	0.782	0.819	0.74	0.896	0.882	0.812	0.93	Temporal cohesion, tense and aspect repetition, mean
			Syr	ntactic Co	omplexity	7					
67	SYNLE	SYNLE	4.019	4.054	6.068	5.698	3.413	3.928	5.157	5.155	Left embeddedness, words before main verb, mean
68	SYNNP	SYNNP	0.873	0.914	0.831	0.962	0.916	1.092	0.883	0.823	Number of modifiers per noun phrase, mean
69	SYNMEDpos	MEDwtm	0.684	0.643	0.656	0.661	0.758	0.719	0.67	0.68	Minimal Edit Distance, part of speech
70	SYNMEDwrd	MEDawm	0.917	0.91	0.899	0.9	0.932	0.928	0.919	0.927	Minimal Edit Distance, all words
71	SYNMEDlem	MEDalm	0.895	0.881	0.884	0.887	0.921	0.916	0.905	0.915	Minimal Edit Distance, lemmas
<mark>72</mark>	SYNSTRUTa	STRUTa	0.061	0.084	0.076	0.079	0.067	0.054	0.077	0.069	Sentence syntax similarity, adjacent sentences, mean
<mark>73</mark>	SYNSTRUTt	STRUTt	0.056	0.077	0.063	0.058	0.07	0.053	0.068	0.046	Sentence syntax similarity, all combinations, across paragraphs, mean

No.	Label	Label V2.x	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Full description				
	Syntactic Pattern Density														
74	DRNP	n/a	390	345.669	377.115	380.908	374.408	331.874	383.604	352.995	Noun phrase density, incidence				
75	DRVP	n/a	186	221.194	196.616	176.089	222.072	200.525	199.536	215.064	Verb phrase density, incidence				
76	DRAP	n/a	27	39.529	22.562	38.925	35.207	45.534	25.522	38.113	Adverbial phrase density, incidence				
77	DRPP	n/a	125	98.402	124.899	132.53	98.849	99.825	139.985	100.726	Preposition phrase density, incidence				
78	DRPVAL	AGLSPSVi	5	13.457	4.835	4.634	8.802	3.503	9.281	5.445	Agentless passive voice density, incidence				
79	DRNEG	DENNEGi	10	9.251	7.252	4.634	4.739	3.503	7.734	9.982	Negation density, incidence				
80	DRGERUND	GERUNDi	19	15.98	25.786	9.268	20.311	24.518	20.108	27.223	Gerund density, incidence				
81	DRINF	INFi	16	21.867	16.116	15.755	14.218	25.394	11.601	24.501	Infinitive density, incidence				
			Wo	rd Inform	ation										
82	WRDNOUN	NOUNi	259	249.79	271.555	300.277	278.944	291.593	280.742	254.991	Noun incidence				
83	WRDVERB	VERBi	119	116.905	119.258	103.8	134.733	112.083	111.368	131.577	Verb incidence				
84	WRDADJ	ADJi	97	125.315	99.919	108.433	76.506	98.074	85.847	103.448	Adjective incidence				
85	WRDADV	ADVi	51	68.965	41.902	56.534	54.164	74.431	44.083	70.78	Adverb incidence				
86	WRDPRO	DENPRPi	63	42.893	62.853	26.877	50.779	42.032	51.817	64.428	Pronoun incidence				

No.	Label	Label V2.x	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Full description
87	WRDPRP1s	n/a	0	0	1.612	0	0	0	7.734	0	First person singular pronoun incidence
88	WRDPRP1p	n/a	20	9.251	18.533	0.927	10.156	3.503	19.335	0.907	First person plural pronoun incidence
89	WRDPRP2	PRO2i	2	1.682	0	0	10.833	0	0	1.815	Second person pronoun incidence
90	WRDPRP3s	n/a	8	5.046	8.864	1.854	10.156	14.886	9.281	0	Third person singular pronoun incidence
91	WRDPRP3p	n/a	21	20.185	26.591	13.902	12.187	13.135	9.281	52.632	Third person plural pronoun incidence
<mark>92</mark>	WRDFRQc	FRCLacwm	2.169	2.145	2.205	2.197	2.25	2.144	2.174	2.25	CELEX word frequency for content words, mean
93	WRDFRQa	FRCLaewm	2.99	2.861	2.969	2.982	3.014	2.904	2.981	2.905	CELEX Log frequency for all words, mean
<mark>94</mark>	WRDFRQmc	FRCLmcsm	1.081	1.365	0.885	1.46	1.387	1.421	1.05	1.141	CELEX Log minimum frequency for content words, mean
95	WRDAOAc	WRDAacwm	344.76	380.108	363.028	405.358	330.52	376.264	361.5	328.388	Age of acquisition for content words, mean
96	WRDFAMc	WRDFacwm	573.48	569.512	574.074	563.047	577.312	572.274	568.535	578.787	Familiarity for content words, mean

No.	Label	Label V2.x	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Full description
97	WRDCNCc	WRDCacwm	362.573	359.662	365.104	352.62	386.566	361.899	374.944	381.598	Concreteness for content words, mean
98	WRDIMGc	WRDIacwm	395.863	391.068	399.204	393.545	411.805	394.8	407.059	416.067	Imagability for content words, mean
99	WRDMEAc	WRDMacwm	429.913	420.704	438.61	418.976	434.268	419.056	433.841	449.465	Meaningfulness, Colorado norms, content words, mean
100	WRDPOLc	POLm	3.71	3.637	3.615	3.965	4.103	3.871	3.742	3.998	Polysemy for content words, mean
101	WRDHYPn	HYNOUNaw	6.639	6.726	6.13	6.091	5.953	6.587	6.503	6.978	Hypernymy for nouns, mean
102	WRDHYPv	HYVERBaw	1.753	1.581	1.632	1.538	1.661	1.468	1.523	1.635	Hypernymy for verbs, mean
103	WRDHYPnv	HYPm	1.828	1.822	1.757	1.883	1.744	1.935	1.902	1.907	Hypernymy for nouns and verbs, mean
			Rea	dability							
104	RDFRE	READFRE	50.886	37.658	43.442	42.182	60.002	52.561	52.812	54.305	Flesch Reading Ease
105	RDFKGL	READFKGL	10.849	13.281	12.425	12.433	8.878	10.04	10.482	10.405	Flesch-Kincaid Grade level
106	RDL2	L2	10.114	11.039	13.023	12.504	11.5	8.336	11.14	12.506	Coh-Metrix L2 Readability

\*the green highlighted ones refer to selected indices