Crossley, S. A., Tywoniw, R., & Choi, J. S. (in press). The Tool for Automatic Measurement of Morphological Information (TAMMI). *Behavior Research Methods.*

Title: The Tool for Automatic Measurement of Morphological Information (TAMMI)

Authors: Scott A. Crossley, Rurik Tywoniw, and Joon Suh Choi

Abstract: This study documents and assesses the the Tool for Automatic Measurement of Morphological Information (TAMMI). TAMMI calculates measures related to basic morpheme counts, morphological variety, morphological complexity, morpheme type-token counts, and variables found in the MorphoLex database (Sánchez-Gutiérrez et al., 2017) including morpheme frequency/length, morpheme family size counts and frequency, and morpheme hapax counts. These measures are assessed in two studies that include a word frequency measure as a control variable. The first study examined links between morphological variables and judgements of reading ease in a corpus of ~5,000 reading excerpts, finding that variables related to derivational variety, word frequecy, affix frequency, and morpheme counts explained 40% of the variance in the reading scores. The second examined links between morphological variables and human assessments of vocabulary proficiency in a corpus of ~7,000 essays written by English language learners (ELLs), finding that the number of morphemes, morpheme variety, and the number of roots explained 21% of the variance in the human assessments.

An important component of reading in English is decoding, wherein strings of characters are segmented to identify words and assign meaning to those words. A confound in decoding is that many frequent words in English contain morphemes that may make decoding more difficult because the reader needs to parse not just the root meaning of the word but also the inflectional and derivational morphemes that may be attached to it. Morphologically complex words, thus, are an important predictor of comprehension difficulty for students with reading difficulties. (Berninger et al., 2003; Nagy & Anderson, 1984).

Morphemes are not only related to reading processes, but also to language production. Acquiring morphological knowledge is an important part of learning both a first and second language (DeKeyser, 2005; Lardiere, 2006). While first language (L1) learners eventually acquire native proficiency in morpheme use, second language (L2) learners often do not. Even highly proficient L2 learners show difficulties in morpheme production in terms of morphological commissions, omissions, and substitutions (Larsen-Freeman, 2010; Todeva, 2010). L2 learners also show variability in their morpheme production such that morphemes present in one phrase may be absent in the next phrase even though they are expected (Long, 2003). Research into L2 morpheme production has demonstrated that inflectional morpheme knowledge develops first followed by the use and increased accuracy of derivational morphemes, which takes longer to develop (Green et al., 2003).

Recent studies have indicated that words with derivational morphemes make up around 30% of the words found in the most common 3,000-word families in English. The most common word families in English (i.e., the 1,000 most common word families) include almost 5,000 lemmas. As an example, the word family for *avoid* includes *avoid, avoidance, avoidable*, and *unavoidable* plus all related inflections. Thus, to know a word family, a reader or speaker needs to know at least five time as many lemmas (Laufer & Cobb, 2020) as they do root words. In total, it is estimated that 60% of new words encountered in a text are morphologically complex (i.e., they contain at least a prefix or a suffix, Angelelli, Marinelli, & Burani, 2014).

Thus, the ability to accurately measure the morphological complexity of words is an important component of understanding word and morpheme processing and the effects that this processing has on comprehension and production. Currently, there are limited approaches to measuring morphological complexity in texts. There is at least one available database that tallies root and affix appearances in English (MorphoLex, Sánchez-Gutiérrez et al., 2017), but the database does not include an interface that allows for automatic calculations of morpheme counts. There is also a website (https://www.lextutor.ca/cgi-bin/morpho/lex/) that will take a single text as input and output all the inflectional and derivational affixes for the word families in that text (confusingly, also called MorphoLex, Laufer & Cobb, 2020), but it also does not report morpheme complexity by text or word. To our knowledge no tool exists that will automatically calculate the complexity of various types of morphemes in a text. Such a tool would allow researchers, practitioners, and material developers to assess the difficulty of texts in terms of morphological complexity or single words to be used in psychological studies.

The goal of this paper is to document and assess the mature version of the open-source Tool for Automatic Measurement of Morphological Information (TAMMI 2.0). An initial version of the tool (TAMMI 1.0) was introduced in Tywoniw and Crossley (2020) but was never publicly released. TAMMI 1.0 calculated the number of tokens in a text that contained inflection and derivational morphemes along with the number of tokens that did not contain inflection and derivational morphemes. TAMMI 2.0 includes measures related to basic morpheme counts, morphological variety, morphological complexity, morpheme type-token counts, and variables found in the MorphoLex database (Sánchez-Gutiérrez et al., 2017) including morpheme frequency/length, morpheme family size counts and frequency, and morpheme hapax counts.

We assess the variables reported in TAMMI 2.0 in two studies. The first study examines links between morphological variables and judgements of reading ease in a corpus of ~5,000 reading excerpts. The underlying hypothesis is that texts that are easier to read will be less morphologically complex. The second study investigates links between morphological variables and human assessments of vocabulary proficiency in a corpus of ~7,000 essays written by English language learners (ELLs). The hypothesis tested in the second study is that more advanced ELLs will produce words that are more morphologically complex.

Morphology

Morphology is the study of word structures and how those structures interact with phonology, syntax and semantics. Words have phonological properties that combine to comprise meaningful parts (the morphemes, Dell, 1986; Shattuck-Hufnagel, 1983; Spencer & Zwicky, 2017). Morphemes have stable meaning across words and cannot be divided into smaller units without changing meaning. Free morphemes can stand on their own (e.g., *kick, sad*). Bound morphemes cannot stand alone, and they also can be root morphemes (e.g., *mort,* which refers to life or death, but cannot be used as English word in isolation) or affixes (i.e., morphemes that attach to free or root morphemes like the -*al* in *mortal* and the -*ing* in *kicking*). Prefixes are a type of affix that come before free or bound root morphemes (e.g., the *un-* in *unhappy*) while suffixes are a type of affix that come after free or bound root morphemes (e.g., the -*ly* in *sadly*). Affixes can be inflectional, which serve grammatical functions (e.g., the -*s* in *He kicks* which is used in third person singular verbs) and do not change the part of speech or underlying meaning of the word to which they attach. Affixes can also be derivational, which alter the meaning or the part of speech of words (e.g., the -*ment* in *government* which changes a verb into a noun or the *de*- in *demystify* which changes the meaning of the word to its antonym). Words with derivational affixes are often referred to as derived words (Clahsen et al., 2010). Derivations are further distinguished from inflections because derivational processes are not as exact as inflectional patterns (Bauer, 2008), following probabilistic distributions rather than deterministic rules (Booij, 2010). For example, the nominalizer -*ment* in *government* can be applied to *commit* for *commitment*, but less so to *approve* for *approvement* (where *approval* and *approbation* are considered more acceptable nominalizations). Derivational affixes can also vary in their productivity, or the likelihood of encountering or using the affix in a novel word. It is thus argued that derivations require a nuanced understanding of derivation construction and the adding of new entries into the mental lexicon. This is different from the use of inflectional morphemes which follow more deterministic structures (Booij, 2010).

Both inflectional and derivational morphological features of language can provide important avenues to understand and study the implicit knowledge learners have about a language. A prime example of this is the Wug Test, which was used to assess the unconscious awareness of morphological building blocks in words as part of child language development (Berko, 1958). Morphological awareness is also an important component of understanding literacy development (Carlisle and Feldman, 1995) and plays an important role in developing word reading and text comprehension (Angelelli, Marinelli, & Burani, 2014; Deacon & Kirby, 2004). As readers are exposed to a greater number of words with affixes, reading success comes to depend on the ability to process morphologically complex words efficiently and accurately, which depends on the morphological knowledge of the reader (Carlisle, 2000; Singson, Mahoney, & Mann, 2000).

For instance, Nagy, Anderson, Schommer, Scott, and Stallman (1989) reported that derived words, the number of words in the word family, and the frequency of words in the word family were associated with word recognition speeds such that words with less morphologically complex words were recognized more quickly. Carlisle and Katz (2006) examined word reading in relation to inflectional and derivational morphemes, base word frequencies, word family size and frequency, and word length. They used a principal component analysis to combine these variables into two components related to morpheme constitution and word family exposure and found that both components explained significant variance in word reading. They also found that older readers and better readers performed better in reading words with inflectional and derivational morphemes. Lastly, Amenta and Crepaldi (2012) reported that reading difficulty may result from the type of morphemes attached to a word, noting that while inflectional morphemes are limited in number and semantics, derivational morphemes are more complex and may influence word decoding to a stronger degree.

Morpheme knowledge is also an important predictor of second language (L2) proficiency, and the study of morpheme acquisition orders in L2 learners likely led to the establishment of second language acquisition as a field of research (Larsen-Freeman, 2010). Most L2 studies have focused on the acquisition of inflectional morphemes that serve grammatical roles in English, which cause specific difficulties for L2 learners (especially adults; Clahsen et al., 2010; MacWhinney, 2005; Murakami & Ellis, 2022). Research has demonstrated that L2 learners often omit inflectional morphemes or substitute morphemes (Larsen-Freeman, 2010; White, 2003). The reasons for the unsystematic use of inflectional morphemes in L2 learners is unclear, but studies indicate that the perceptual salience, semantic complexity, morphophonological regularity, syntactic category, and frequency of inflectional morphemes can help explain acquisition difficulties (Goldschneider & DeKeyser, 2005). Examining derived words in L2 learners is much less common (Clahsen et al., 2010). The few studies available indicate that L2 learners may process derived words similarly regardless of their first language (Koda, 2000) and that there are no priming effects for inflected morphemes and reduced priming for derivational morphemes, while priming effects are found in native speakers for both inflectional and derivational morphemes (Silva & Clahsen, 2008).

Measuring Morphology

There have been numerous studies that have quantified the number and types of morphemes found in texts to better understand the role morphemes play in language. For instance, Laws & Ryder (2014) generated frequency norms for derivational affixes in the British National Corpus (BNC Consortium, 2007) using a list of attested English affixes (Stein, 2007). Their final database included lists of word types that had specific derivational affixes and the number of tokens in the ten-million word BNC spoken component that contained these affixes. They found that about ten percent of word tokens included derivational affixes and that there were slightly more prefixes in the BNC than suffixes. They also reported that a higher proportion of word types involved suffixes.

Sánchez-Gutiérrez et al. (2018) developed and tested the MorphoLex database. The database includes morphological information for the 68,624 words found in the English Lexicon Project (ELP, Balota et al., 2007). Sánchez-Gutiérrez et al. initially segmented each word into prefixes, suffixes, and root morpheme using automated techniques. They then manually performed a series of changes to the automated annotations to remove inflectional morphemes, contractions, and normalize the treatment of neoclassical compounds (compounds that comprise at least one lexeme from the classical languages). They then selected roots for each base morpheme and then revised annotations to control for allomorphy, to eliminate remaining pseudo-derivations, and correct mis-annotated affixes and roots.

The final version of the MorphoLex database reports the root and any prefixes and suffixes for each word. The database also reports six morpheme variables for affixes and three for roots related to morphological family size, morphological frequency, affix productivity, affix length, and the percentage of other words in the morpheme family that are more frequent. Sánchez-Gutiérrez et al. tested the database on lexical decision latencies for 4,724 morphologically complex nouns (i.e., nouns with one root and suffix) reported in the ELP. They found that words with higher root frequency and shorter suffix length lead to faster lexical decision latencies.

Studies have also examined morphological production in learner data. For example, Lüdeling, Hirschmann, & Shadrova (2017) examined English learners' acquisition of morphologically complex verb forms. They examined German learners' morpheme errors when producing morphologically complex verbs and found that while learners used verbal morphology productively, they did not demonstrate the productivity of native speakers. Brezina and Pallotti (2019) developed a measure of complexity called the Morphological Complexity Index (MCI), which calculated the average inflectional diversity for the occurrences of a given word class within a text in a manner similar to lexical diversity calculations like type-token ratio. The basic notion of the MCI calculations was that texts with a greater diversity of inflectional morphemes would be more morphologically complex. In their first study, they reported that writing samples produced by non-native speakers of English had lower MCI than samples produced by native speakers and that the effect of lower MCI counts was stronger for L2 learners from lower proficiency levels. In a second study, they found that MCI counts were constant across advanced

L2 learners and native speakers of English, attributing the finding to the relative simplicity of English inflectional morphemes.

Tool for Automatic Measurement of Morphological Information (TAMMI 2.0)

In this study, we introduce and assess TAMMI 2.0. TAMMI 2.0 is a freely available text natural language processing (NLP) tool specifically designed to annotate and count morphological features in texts. The tool is available at linguisticanalysistools.com and the source code for the tool is available at https://github.com/scrosseye/tammi.[1] The downloadable version of TAMMI 2.0 is available for both Mac and Windows operating systems and features a graphical user interface that allows users to batch process texts (see Figure 1).

[Insert Figure 1 here]

In developing TAMMI 2.0, we provide automatic calculations for the MorphoLex data frame provided by Sánchez-Gutiérrez et al. (2017). MorphoLex was assessed for reliability in a series of analyses that examined the influence of MorphoLex variables on the lexical decision latencies of 4,724 morphologically complex nouns. We also included an automatic calculation of morphological complexity index (MCI) based on inflections as detailed by Brezina and Pallotti (2019). In addition, we calculated an MCI for derivational morphemes and developed new morphological complexity indices based on morphological variety and type-token ratios for both inflectional and derivational morphemes. Lastly, we calculate a number of basic morpheme counts. All morpheme indices included in TAMMI measure morphological complexity.

Pre-processing. Before calculating any morpheme features, texts are pre-processed using spaCy's small, English model based on the web (version 3.2.1; Honnibal & Montani, 2017). spaCy is used to first tokenize the texts to extract the words. Next, stop words from spaCy's stop

[1] This script does not include the graphical user interface code.

word list are removed. The stop word list includes 326 stop words and contractions. Stop words include pronouns, copula verbs, conjunctions and connectives, prepositions, some adverbs like *mostly, formerly,* and *really,* and delexical verbs like *give, have, done,* and *be*. Contractions include *'ll, 're,* and *n't*. Lastly, only alphabetic tokens are kept. This removes all numbers and punctuation.

Stop word lists generally require additional tailoring to perform better on different types of texts (Zaman, Matsakis, & Brown, 2011), and general all-purpose stop word lists often have limitations in performance and in rationale. These concerns have resulted in the creation of more than fifty different popular lists of English stop words (Nothman, 2013). We have elected to use spaCy's default stop word list for efficiency and consistency. The default list is efficient because it is designed to be compatible with spaCy's NLP pipeline. It also produces similar results with multiple other popular packages because they share the same origin (a stop word list derived by Glasgow Information Retrieval Group[2] and modified by Stone, Dennis, & Kwantes, 2010).

Basic morpheme counts. TAMMI 2.0 includes basic morpheme counts for inflections and derivational morphemes. The inflections are counted using spaCy by assessing the differences in the number of characters between each token and its lemma. spaCy uses rule-based approaches for English that include part of speech (POS) tagging to assign base forms (i.e., uninflected forms) to tokens. For instance, *swimming* used as a noun (i.e., gerund) would not be lemmatized but *swimming* as a verb would be lemmatized to *swim*. Similar results would hold for *boring* used as a participial adjective, which would not be lemmatized, and *boring* or *bored* used as a verb, which would both be lemmatized to *bore*. The precision of lemmatization depends on the

[2] http://ir.dcs.gla.ac.uk/resources/linguistic_utils/stop_words

spaCy POS tagger, which reports an accuracy of ~97%[3] for the larger models when evaluated against the OntoNotes 5.0 corpus (Weischedel et al., 2013); however, this accuracy is likely lower for smaller models and for data on which the tagger was not trained.

Derivational morphemes are calculated using MorphoLex, which provides counts for prefixes, suffixes, and affixes as well as the number of compound words (i.e., words that have more than one root morpheme). In addition, TAMMI calculates the average number of morphemes per word. All basic morpheme counts are normed by the number of content words in the text (i.e., all words that are not in the spaCy stop word list). TAMMI 2.0 also computes normed indices by taking the count for each variable and dividing it by the number of content words with the relevant morpheme. For example, when calculating the number of suffixes, TAMMI 2.0 will sum the number of suffixes in the text and norm that sum by the number of words that contain a suffix. The same is done for affixes and total suffixes.[4] Features normed by relevant morpheme type will likely not perform well on texts with simple morpheme use because each word may only contain a single morpheme type (i.e., a text may contain only a single suffix in all words that contain a suffix, giving a normed score by suffix of 1). Thus, users should only use the relevant normed morpheme counts when examining longer texts that are representative of more advanced language use.

Morphological variety. The inflection morphological variety feature in TAMMI 2.0 is based on a within-subset variety score reported in Brezina and Pallotti (2019) in which content words from each text are broken into windows of 10 words (plus a window of 1-to-9 for any remaining content words at the end of the text). Inflectional morpheme types (e.g., -*s* and -*ed*) for each

[3] Reliability for spaCy is reported at https://github.com/explosion/spacy-models/blob/master/meta/en_core_web_sm-3.2.0.json.
[4] A worked example for calculating Morpholex scores based on a single sentence is available at https://github.com/scrosseye/Tammi-Analyses

content word in the window, and null tokens for words without inflections in the window, are counted for each 10-word window. This count is then divided by the total number of windows in the text to calculate within-subset variety scores for the entire text. This improves upon simply counting the number of inflectional types in the text because a simple type count would likely correlate with text length. A similar approach is used to assess derivational morpheme variety. However, since a content word could have multiple derivational morphemes, the windows of 10 words and/or null counts could have multiple derivational morphemes per word. Thus, a window of ten derivational morphemes and/or null counts may reflect 10 content words or fewer.

Morphological complexity. TAMMI 2.0 calculates an index for inflectional morphemes based on the MCI reported in Brezina and Pallotti (2019) by using the morphological variety counts above. For inflections, a between-subset diversity score is calculated. The between-subset diversity score is the average number of unique morphemes when comparing subsets where subset are the same 10-word windows found in the within-subset variety score. As an example of unique morphemes, *I loved him* and *she loves him* each have one unique morpheme, *-ed* and *-s*. The within-subset variety score is then divided by the between-subset diversity score (i.e., this score is then divided by the number of subsets minus 1). The same approach is followed to produce an MCI for the derivational morphemes, which was not reported by Brezina and Pallotti (2019).

Morpheme type-token counts. TAMMI 2.0 includes indices of type-token ratios (TTR) for both inflectional and derivational morphemes. For inflectional morphemes, we use the number of unique inflectional morphemes by 10 content word window divided by the length of the window (knowing that the last window may be less than 10 words) and average the score across the text.

For derivational morphemes, we calculate a similar metric, but we use a 10-morpheme window because some content words have more than one derivational morpheme.

Frequency, family size, and hapax counts. TAMMI 2.0 depends on MorphoLex to calculate variables related to frequency/length, family size counts and frequency, and hapax counts. TAMMI 2.0 matches tokens reported in spaCy to the MorphoLex dictionary. Like basic counts, TAMMI 2.0 computes mean scores for MorphoLex frequency/length, family size counts and frequency variables within a text by taking the count for each variable and dividing it by the number of content words (i.e., all words not in the spaCy stopword list) in the text to provide a normed score. Additionally, there are normed scores by the number of prefixes, suffixes, and total affixes. The MorphoLex variables calculated in TAMMI 2.0 are discussed below.

Morpheme frequency/length counts. For roots, prefixes, suffixes, and all affixes, TAMMI 2.0 extracts frequency counts for morphemes from MorphoLex. The frequency count comes from the HAL counts found in the ELP (Balota et al., 2007). TAMMI 2.0 computes a raw frequency count and a logged frequency count. TAMMI 2.0 also calculates the average length of the roots, prefixes, suffixes, and all affixes.

Morpheme family size counts. For roots, prefixes, suffixes, and all affixes, TAMMI 2.0 derives family size counts from MorphoLex. Family sizes for morphemes are calculated by counting the number of word types to which a morpheme can attach itself. As an example, in the pool of words *attendance, pleasance, pleasure, appearance*, the suffix -*ance* has a family size of 3, but the root *pleas* has a family size of 2 (example taken from Sánchez-Gutiérrez et al., 2017). For roots, family size is calculated by the number of words a root can produce (e.g., the count for the number of words that have *theo* as a root).

Morpheme family size frequency. TAMMI 2.0 also reports the percentage of other words in the family that are more frequent (PFMF) from MorphoLex. This feature counts the percentage of morphemes per word that are more frequent by dividing the number of more frequent words in a family by the total number of family members. For instance, *word, wordiness,* and *wordlessly* all have the same root (i.e., *word*), but the word *word* is the most frequent type in the family (PFMF = 0%) whereas *wordlessly* has 10 terms that are more frequent (PFMF = 45%) and *wordiness* has 15 types that are more frequent (PFMF = 70%; example taken from Sánchez-Gutiérrez et al., 2017). Thus, a lower value indicates a word that is more frequent in the family, and higher PFMF values can contribute to greater morphological complexity.

Hapax counts. Hapaxes are defined in MorphoLex as words or roots that only appear once in a corpus. Affixes that attach to a greater number of hapaxes are more productive and can be used to create new words. TAMMI 2.0 derives two types of hapax counts from MorphoLex: the number of prefixes/suffixes/affixes that are attached to hapaxes and the number of hapaxes that include prefixes/suffixes/affixes.

Current Study

We report on two assessments of TAMMI 2.0 that provide case studies and validation for the tool. The first study assesses the strength of TAMMI 2.0 indices to predict text readability. The readability of a text can be influenced by the morphological complexity of words contained with that text because a certain level of morphological awareness is required to fluently decode morphologically complex words (Nagy, Berninger, & Abbott, 2006). Words that are morphologically complex should thus lead to lower reading comprehension. The second study examines the strength of TAMMI 2.0 indices to predict the language proficiency of second language (L2) writers of English. More advanced learners of English would have better developed morphological skills (Zhang & Koda, 2012) which should be registered in their written production of the L2. For each study, we include a word frequency measure as a control variable to examine if the TAMMI variables explain variance beyond a standard measure of lexical sophistication.

Study 1

Methods

Corpus. To assess the links between morphological elements of texts and text readability, we used the CommonLit Ease of Readability (CLEAR) corpus (Crossley, Heintz, Choi, Batchelor, & Karimi, 2021; Crossley, Heintz, Choi, Batchelor, Karimi, & Malatinszky, 2023). The CLEAR corpus contains 4,724 text samples totaling ~800,000 words. The corpus was developed to model and test various readability metrics. To collect unique readability scores for each excerpt, teachers were recruited via email from CommonLit's internal teacher pool. Teachers were shown two text excerpts at a time and instructed to select which excerpt they believed was easier to read for students. After removing outliers, data was kept from 1,116 teachers, who made 111,347 overall comparison judgments. A Bradley-Terry model (Bradley & Terry, 1952) was used to compute pairwise comparison scores for the teachers' judgments of text ease to calculate unique readability scores for each excerpt. The final scores reflect the "Easiness" in terms of comprehension for each excerpt in the corpus.

Frequency Measure. We included a measure of word frequency as a control variable. Word frequency measures calculate the number of times a word occurs in a text corpus. Words that are less frequent are considered more sophisticated than words that are more frequent. Studies have indicated that word frequency is a strong predictor of text readability (Crossley et al., 2023) and of language proficiency (Kyle & Crossley, 2015). We used the Tool for the Automatic Analysis of Lexical Sophistication (TAALES; Kyle, Crossley, & Berger, 2018) to calculate a logged word frequency measure derived from the news section of the Corpus of Contemporary American English (COCA; Davies, 2009).

Statistical Analyses. To predict text reading ease scores found in the CLEAR corpus, we used the morpheme indices normed by content words calculated from TAMMI 2.0 along with the frequency measure as predictor variables in a linear model. We did not use morpheme indices normed by relevant morpheme counts. We first ensured that none of the TAMMI 2.0 variables or the frequency measure correlated strongly with text length (r > .699). All variables correlated at r < .010 with text length except MCI for inflections, which correlated at r = -0.200. Considering this was a low correlation, we included all the TAMMI 2.0 variable in the analyses. We next calculated bivariate Pearson correlations for all TAMMI 2.0 variables and the word frequency measure using the cor.test() function in R (R Core Team, 2020) to identify highly collinear features among the morpheme variables. If two or more variables correlated at r > .699, the variable(s) with the lowest correlation with the ease of readability score was removed and the variable with the higher correlation was retained. We also only retained variables that demonstrated at least a small relationship with the ease of readability scores (r > .099).

We used the CARET package (Kuhn, 2008) in R to develop linear models. Model training and evaluation were performed using a ten-fold cross-validation model using stepwise selection from the leapSeq() function. Estimates of accuracy are reported using the amount of variance explained by the developed models (R^2). The model was checked for suppression effects. The relative importance of the indices in each model was calculated using the calc.relimp() function in the relaimpo package (Grömping, 2006) using the lmg metric (Lindeman, Merenda, & Gold, 1980). lmg takes into account both the direct relationship between the independent and dependent variable (i.e., the bivariate correlation) and the indirect relationship between the independent and dependent variable (i.e., the amount of variance explained when included in a multivariate model).[5]

Results

Correlations. Of the 42 TAMMI variables assessed, 32 variables were removed because of multi-collinearity or because they did not report at least a small effect size with the readability score. All correlations between the TAMMI variables and the frequency measure were below $r <$ .50. Correlations among the remaining variables indicated small to medium relationships ($r <$ .10 and > .50). The strongest correlation was for derivational MCI. The weakest correlation was for inflectional MCI. Correlations are reported in Figure 2.

[Insert Figure 2 here]

Linear model. The 11 variables that remained after controlling for multicollinearity and that showed at least a weak relationship with reading ease scores were entered into a linear model along with the word frequency measure. The ten-fold cross validation model using step-wise selection included nine variables and reported $r = .633$, $R^2 = .401$, $F(9, 4714) = 393.9$, $p < .001$ (see model parameters summarized in Table 1). The relative importance metrics indicate that the strongest predictors of reading ease were related to diversity of derivational morphemes, word frequency, family frequency (suffixes), and derivational TTR. Post-hoc tests indicated all linear model assumptions were met.

[Insert Table 1 here]

Discussion

[5] All R scripts and data for the analyses in this paper are available at https://github.com/scrosseye/Tammi-Analyses

Our first study examined the potential for morpheme counts calculated by TAMMI 2.0 to predict text readability scores in the CLEAR corpus. Our hypothesis was the morpheme counts related to complexity would be related to difficulty in text decoding such that reading excerpts that contain a greater number of words with more complex morpheme structures would be judged to be more difficult to understand (Carlisle, 2000; Singson, Mahoney, & Mann, 2000). Our correlational analysis and linear model strongly support this notion indicating that, after controlling for multicollinearity and small effect sizes, 10 TAMMI 2.0 variables showed small to medium effects with text readability scores. Seven of these variables were significant predictors in a linear model that explained almost 40% of the variance in the readability scores.

The linear model indicated that the strongest TAMMI predictor of text readability was the morphological complexity index (MCI) that we developed for derivational morphemes. The co-efficient estimate indicated that texts that were more difficult to read included a greater diversity of derivational morphemes than text that were easier to read. Similarly, TTR counts for derivational morphemes, which also measure morpheme diversity was also a significant, and negative, predictor. Derivational morphemes have a greater number of types and are more productive (i.e., they can function as both prefixes and affixes, can change a word's meaning and part of speech, and can combine to make multi-morphemic words). Excerpts with a greater number of of derivational morphemes are, thus, more difficult to process (Amenta & Crepaldi, 2012).

Our word frequency measure derived from COCA was also a strong predictor of text readability. It was the second strongest predictor after the derivational MCI variable and, as expected, indicated that excerpts that were more difficult to read contained more infrequent words. Normed morpheme frequency counts were also an important predictor of readability. In terms of frequency, higher suffix and prefix family frequency led to lower comprehension scores. Higher root frequency and prefix frequency led to increased comprehension scores. Similar results were reported for normed Morpheme family size frequency. This finding supports the findings of Sánchez-Gutiérrez et al. (2018) who reported that word frequency at the word and morphological levels impacts word processing and text readability.

Lastly, texts that contained a greater number of inflections in general were predicted to be easier to read. There is a small number of inflectional morpheme types (only eight in English) and these morphemes provide grammatical structure to text. The structural elements of inflection morphemes may lead to texts that are easier to read. Additionally, inflectional morphemes may be more common in literary texts that are more descriptive (e.g., easier to read texts may contain more comparatives, superlatives, and possessives) than more information-dense and less linear text types. The increased number of descriptives in narrative texts may be lead to increased inflectional morphemes.

Overall, the results provide a measure of validation for TAMMI 2.0 in that the morphological features it measures are predictive of text readability following theoretical expectations related to how word decoding can be made more difficult by the complexity of a word. Additionally, the results provide some evidence about how morpheme complexity and text readability interact, providing indications for the role that morphemes may play in text comprehension.

Study 2

English Language Learning Insight, Proficiency and Skills Evaluation (ELLIPSE) corpus. The ELLIPSE corpus comprises 6,482 essays written by ELLs. All essays were written during state-wide standardized annual testing in the United States. The essays were written on 29 different independent prompts that required no background knowledge on the part of the writer. Each essay was scored by two normed human raters on a five-point rubric for English language proficiency including an overall score of English proficiency and analytic scores for cohesion, syntax, vocabulary, phraseology, grammar, and conventions. Reliability for the proficiency scores as reported by a Many-Facet Rasch Measurement (MFRM) was high in terms of texts, raters, and scales (Crossley et al., in press). For the purposes of this study, we focus on the overall score of English proficiency. A high score indicated native-like facility in the use of language including syntactic variety, word and phrase choice, text organization, grammar, and convention. A low score indicated a limited range in the above criteria.

Statistical Analysis. The statistical analysis used for the ELLIPSE corpus was similar to that used in the CLEAR corpus. To predict proficiency score, we used the morpheme indices normed by content words calculated from TAMMI 2.0 as predictor variables in a linear model. We did not use morpheme indices normed by relevant morpheme counts. We included the same frequency measure used in Study One as a control variable (COCA news frequency). We ensured that none of the variables correlated strongly with text length. Like the CLEAR corpus analysis, all variables correlated at $r < .010$ with text length except MCI for inflections, which correlated at $r = -0.536$ and MCI for derivational morphemes which correlated at $r = -0.395$. Considering these were relatively high correlations, the two MCI variables were removed from the subsequent analyses. We also calculated bivariate Pearson correlations for the remaining TAMMI 2.0 variables using the cor.test() function in R (R Core Team, 2020) to identify highly collinear features. If two or more variables correlated at $r > .699$, the variable(s) with the lowest correlation with proficiency score was removed and the variable with the higher correlation was retained. We also only retained variables that demonstrated at least a small relationship with the proficiency score (r > .099).

We used the CARET package (Kuhn, 2008) in R to develop a linear model. Model training and evaluation were performed using a ten-fold cross-validation model using stepwise selection from the leapSeq function. Estimates of accuracy are reported using the amount of variance explained by the developed models (R^2). The model was checked for suppression effects. The relative importance of the indices in each model was calculated using the calc.relimp() function in the relaimpo package (Grömping, 2006) using the lmg metric (Lindeman, Merenda, & Gold, 1980).

Results

Correlations. Of the 41 variables assessed, 31 variables were removed because of multi-collinearity or because they did not report at least a small effect size with the proficiency score. Correlations among the remaining TAMMI variables indicated small to medium relationships (*r* < .10 and > .50). The correlation between the proficiency score and the word frequency variable reported *r* = -.179. Correlations are reported in Figure 3.

[Insert Figure 3 here]

Linear model. The ten variables that remained after controlling for multicollinearity and that showed at least a weak relationship with reading ease scores were entered into a linear model along with the word frequency measure. The ten-fold cross validation model using step-wise selection included five variables after removing two variables that reported suppression effects. The model reported r = .463, R^2 = .214, F (5, 6476) = 352.3, p < .001 (see model parameters summarized in Table 2). The relative importance metrics indicate that the strongest predictors of ELL proficiency were related to a greater average use of morphemes, a greater diversity of both inflectional and derivational morphemes, and a greater average number of root morphemes. Word frequency was a significant predictor but demonstrated the lowest relative importance. Post-hoc tests indicated all linear model assumptions were met.

[Insert Table 2 here]

Discussion

Our second analysis examined how language proficiency in ELLs can be predicted based on morphological production within essays written by ELLs. The purpose of this study was to examine links between overall language proficiency and morphological production. Our hypothesis was that less proficient ELLs would produce fewer complex morphemes as a result of difficulties in morphological acquisition, which is difficult for L2 learners (Larsen-Freeman, 2010; Long, 2003; Todeva, 2010). To do so, we used TAMMI 2.0 to calculate morpheme related features in the ~7,000 ELL essays that had been scored for language proficiency by expert raters in the ELLIPSE corpus. Our correlational analysis and linear model supported the notion that morpheme production in ELL writing samples was strongly related to human ratings of language proficiency. After controlling for multicollinearity and small effect sizes, ten TAMMI 2.0 variables showed small to medium effects with language proficiency scores. Four of these variables were significant predictors in a linear model that predicted over 20% of the variance in the language proficiency scores.

In the linear model, the strongest predictor of language proficiency was the average number of morphemes per word. The coefficients indicated that writers who produced more morphemes per word were judged to have stronger language proficiency. This makes sense in terms of previous studies that have shown that L2 learners have difficulty in morpheme production, especially when compared to native speakers (Larsen-Freeman, 2010; Todeva, 2010).

The second strongest predictor was a measure of inflectional variety such that ELL writers that produced a greater variety of inflectional morphemes were judged to have stronger language proficiency. Inflectional morphemes are a closed class of morphemes, so increased use in writing indicates sophistication of the writer while not increasing the overall complexity of the text. Second language acquisition (SLA) studies have demonstrated that adult L2 learners have a difficult time developing accuracy with inflectional morphemes likely because of the grammatical roles they have and the difficulties these roles play with L2 learners (Clahsen et al., 2010; MacWhinney, 2005; Murakamia & Ellis, 2022). These findings indicate that adolescent ELLs may have similar variance in their production accuracy of inflectional morphemes and that the use of inflectional morphology improves the quality of ELL writing rather than complicating it.

The linear model also indicated that more proficient ELLs produce a greater variety of derivational morphemes (as reported in the derivational TTR result) providing additional support that more advanced L2 learners produce a greater variety of morphemes. While derivational morpheme studies are rare in L2 learning, previous studies have indicated that priming effects for derived words are reduced in L2 learners, likely because of weaker lexical networks related to morphemes (Silva & Clahsen, 2008). More advanced learners should have strong networks that allow for the production of a greater number of derived words. Lastly, the linear model results indicated that ELL writers that produced a greater average number of root morphemes (per content word) were also judged to be more proficient. This finding supports previous studies that have reported the use of more complex lexicon in advanced L2 learners (Kyle, Crossley, & Berger, 2018).

Like in our reading corpus, the word frequency measure was a significant predictor of language proficiency. The co-efficients indicate that ELL students who produce texts with more infrequent words are judged to be more proficient. This finding supports numerous previous studies that have shown strong relationships between increased proficiency in L2 learners and their use of more infrequent words.

Overall, these results provide a profile for how morphemes can be used to model language proficiency in ELL writers and demonstrate a use case for TAMMI 2.0 that helps validate the tool. As expected, ELLs that are judged to be more advanced produce a greater number of morphemes overall than those judged to be less advanced. Additionally, more advanced ELL writers produce a greater variety of morphemes, demonstrating stronger knowledge of the morphemic system in English.

Conclusion

The analyses reported in this paper assess the reliability of TAMMI 2.0, which is a mature version of TAMMI 1.0, which was introduced in Tywoniw and Crossley (2020) but never publicly released. TAMMI 2.0 is full open-source and available for free download. TAMMI 2.0 calculates indices related to basic morpheme counts, morphological variety, morphological complexity, morpheme type-token counts, and variables reported by the MorphoLex database (Sánchez-Gutiérrez et al., 2017) including morpheme frequency/length, morpheme family size counts and frequency, and morpheme hapax counts.

TAMMI 2.0 variables were validated in two studies. The first examined links between morphological variables and judgements of reading ease in a corpus of ~5,000 reading excerpts, finding that variables related to derivational variety, affix frequency, and morpheme counts explained 40% of the variance in the reading scores. The second examined links between morphological variables and human assessments of vocabulary proficiency in a corpus of ~7,000 essays written by English language learners (ELLs), finding that the number of morphemes, morpheme variety, and the number of roots explained 21% of the variance in the human assessments. While the findings appear robust, they may be confounded by other language features in the text. For instance, there may be overlap between the morpheme features calculated by TAMMI and measures of word frequency, age of acquisition, lexical decision times, or even word neighborhood counts. We did not control for these potential confounds because our interest was in the morpheme variables alone. However, future studies may want to examine relationships between morphological complexity and lexical sophistication.

The strongest predictors, overall, were the morpheme variety score calculated by TAMMI 2.0. In most cases, these were variety indices developed specifically for TAMMI 2.0 (e.g., derivational MCI, TTR). A number of the variables extracted from MorphoLex were also strong predictors including the morpheme variables related to morpheme and morpheme family frequency. Normed counts unique to TAMMI 2.0 were also predictive including the number of overall morphemes and the number of inflectional morphemes. In general, all variables except the hapax count variables were predictive of either text readability or ELL proficiency. It may be the case that the range of texts sampled in this study did not include advanced enough or specific enough topical domains to elicit rare and novel uses of productive affixes, so measuring hapax was not meaningful.

Like other NLP tools, TAMMI 2.0 is not without limitations. A major limitation is that it cannot assess whether morphemes are used accurately, just that morphemes are used. In some cases, TAMMI 2.0 cannot assess whether the morpheme is accurately attached to the root word (in the case of the number of inflectional morphemes calculated through spaCy). Lastly, TAMMI 2.0 is only available for the English language. Nevertheless, TAMMI 2.0 will allow researchers to examine language at both a granular and larger scale in terms of morpheme use in texts. Two possibilities for TAMMI 2.0 were presented in this paper, but TAMMI 2.0 could be used in a large variety of studies interested in language assessment, language development, language acquisition, text modeling, cognition modeling, and other potential language phenomena that involve morphemes.

Crossley, S. A., Tywoniw, R., & Choi, J. S. (in press). The Tool for Automatic Measurement of Morphological Information (TAMMI). *Behavior Research Methods.*

Open Practices Statement

TAMMI is available at linguisticanalysistools.org and the base code for the tool can be found at https://github.com/scrosseye/tammi. All data and materials for all studies reported in this paper are available at https://github.com/scrosseye/Tammi-Analyses.

Crossley, S. A., Tywoniw, R., & Choi, J. S. (in press). The Tool for Automatic Measurement of Morphological Information (TAMMI). *Behavior Research Methods.*

References

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in psychology, 3*, 232.

Angelelli, P., Marinelli, C. V., & Burani, C. (2014). The effect of morphology on spelling and reading accuracy: a study on Italian children. *Frontiers in psychology, 5*, 1373.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior research methods, 39*, 445-459.

Bauer, L. (2008). Derivational morphology. *Language and linguistics compass*, *2*(1), 196-210.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*(2-3), 150-177.

Berninger, V. B., Nagy, W. E., Carlisle, J. F., Thomson, J., Hoffer, D., Abbott, S., Abbott, R., Richards, T., & Aylward, E. (2003). Effective treatment for children with dyslexia in grades 4–6: Behavioral and brain evidence. In B. Foorman (Ed.), Preventing and remediating reading difficulties: Bringing science to scale (pp. 381–347). Baltimore: York Press.

Booij, G. (2010). Construction morphology. *Language and linguistics compass*, *4*(7), 543-555.

Brezina, V., & Pallotti, G. (2019). Morphological complexity in written L2 texts. *Second Language Research, 35*(1), 99–119.

BNC Consortium. (2007). *British national corpus*. Oxford Text Archive Core Collection.

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and writing, 12*(3), 169-190.

Crossley, S. A., Tywoniw, R., & Choi, J. S. (in press). The Tool for Automatic Measurement of Morphological Information (TAMMI). *Behavior Research Methods.*

Carlisle, J. F., & Feldman, L. B. (1995). Morphological awareness and early reading achievement. In L. Feldman (Ed.). *Morphological aspects of language processing*, pp. 189-209. New Jersey: Lawrence Erlbaum.

Carlisle, J. F., & Katz, L. A. (2006). Effects of word and morpheme familiarity on reading of derived words. *Reading and Writing, 19*, 669-693.

Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language learning, 60*(1), 21-43.

Crossley, S. A., Heintz, A., Choi, J., Batchelor, J., & Karimi, M. (2021). The CommonLit Ease of Readability (CLEAR) Corpus. *Proceedings of the 14th International Conference on Educational Data Mining (EDM)*. Paris, France.

Crossley, S. A., Heintz, A., Choi, J., Batchelor, J., Karimi, M., & Malatinszky, A. (2023). A large-scaled corpus for assessing text readability. *Behavior Research Methods, 55,* 491–507.

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*(2), 159–190. http://doi.org/10.1075/ijcl.14.2.02dav

DeKeyser R. (2005). What makes second-language grammar difficult? A review of issues. *Language Learning, 55*, 1–25.

Deacon, S. H., & Kirby, J. R. (2004). Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. *Applied psycholinguistics, 25*(2), 223-238.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological review, 93*(3), 283.

Crossley, S. A., Tywoniw, R., & Choi, J. S. (in press). The Tool for Automatic Measurement of Morphological Information (TAMMI). *Behavior Research Methods.*

Green, L., McCutchen, D., Schwiebert, C., Quinlan, T., Eva-Wood, A., & Juelis, J. (2003). Morphological Development in Children's Writing. *Journal of educational psychology*, 95(4), 752.

Goldschneider, J. M., & DeKeyser, R. M. (2005). Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A meta-analysis of multiple determinants. *Language learning, 55*(S1), 27-77.

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing

Koda, K. (2000). Cross-linguistic variations in L2 morphological awareness. *Applied psycholinguistics, 21*(3), 297-320.

Kyle, K., & Crossley, S. A. (2015). Automatically Assessing Lexical Sophistication: Indices, Tools, Findings, and Application. *TESOL Quarterly*, *49*(4), 757–786. http://doi.org/10.1002/tesq.194

Kyle, K., Crossley, S. A., & Berger, C. (2018). The Tool for the Automatic Analysis of Lexical Sophistication Version 2.0. *Behavior Research Methods, 50* (3),1030-1046.

Lardiere, D. (2006). Knowledge of derivational morphology in a second language idiolect. In *Proceedings of the 8th generative approaches to second language acquisition conference (GASLA 2006)*. (pp. 72-79). Cascadilla Proceedings Project, Somerville, MA.

Larsen-Freeman, D. (2010). Not so fast: A discussion of L2 morpheme processing and acquisition. *Language Learning, 60*(1), 221-230.

Laufer, B., & Cobb, T. (2020). How much knowledge of derived words is needed for reading?. *Applied Linguistics, 41*(6), 971-998.

Crossley, S. A., Tywoniw, R., & Choi, J. S. (in press). The Tool for Automatic Measurement of Morphological Information (TAMMI). *Behavior Research Methods.*

Laws, J., & Ryder, C. (2014). Getting the measure of derivational morphology in adult speech a corpus analysis using MorphoQuantics. *Language Studies Working Papers, 6*, 3–17.

Long, M. (2003). Stabilization and fossilization in interlanguage development. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 487–535). Malden, MA: Blackwell.

Lüdeling, A., Hirschmann, H., & Shadrova, A. (2017). Linguistic models, acquisition theories, and learner corpora: Morphological productivity in SLA research exemplified by complex verbs in German. *Language Learning, 67*(S1), 96-129.

MacWhinney, B. (2005). A unified model of language acquisition. Handbook of bilingualism: Psycholinguistic approaches, 4967.

Murakami, A., & Ellis, N. C. (2022). Effects of availability, contingency, and formulaicity on the accuracy of English grammatical morphemes in second language writing. Language Learning, 72(4), 899-940.

Nagy, W. E., & Anderson, R. (1984). How many words in printed school English? *Reading Research Quarterly, 19*, 304–330.

Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., & Stallman, A. C. (1989). Morphological Families in the Internal Lexicon. *Reading Research Quarterly, 24*(3), 262–282.

Nagy, W., Berninger, V. W., & Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of educational psychology, 98*(1), 134.

Crossley, S. A., Tywoniw, R., & Choi, J. S. (in press). The Tool for Automatic Measurement of Morphological Information (TAMMI). *Behavior Research Methods.*

Nothman, J., Qin, H., & Yurchak, R. (2018). Stop Word Lists in Free Open-source Software Packages. *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, Ed. Park, E. L., Hagiwara, M., Milajevs, D, and Tan, L. (pp. 7–12). Association for Computational Linguistics, Melbourne, Australia. https://doi.org/10.18653/v1/W18-2502

Sánchez-Gutiérrez, C. H., Mailhot, H., Deacon, S. H., & Wilson, M. A. (2018). MorphoLex: A derivational morphological database for 70,000 English words. *Behavior research methods, 50*, 1568-1580.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In MacNeilage, P. F. (Ed.), *The production of speech* (pp. 109-136). Springer, New York, NY.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and cognition, 11*(2), 245-260.

Singson, M., Mahony, D., & Mann, V. (2000). The relation between reading ability and morphological skills: Evidence from derivational suffixes. *Reading and writing, 12*, 219-252.

Spencer, A., & Zwicky, A.M. (2017). *The handbook of morphology*. Amsterdam, Netherlands: John Benjamins Publishing Company.

Stein, G. (2007). *A dictionary of English affixes: their function and meaning*. Lincom Europa.

Stone, B., Dennis, S., & Kwantes, P. J. (2011). Comparing methods for single paragraph similarity analysis. *Topics in Cognitive Science*, *3*(1), 92-122.

Crossley, S. A., Tywoniw, R., & Choi, J. S. (in press). The Tool for Automatic Measurement of Morphological Information (TAMMI). *Behavior Research Methods.*

Todeva, E. (2010). Multilingualism as a kaleidoscopic experience: The mini universes within. In E. Todeva & J. Cenoz (Eds.), *The multiple realities of multilingualism*. Berlin: Mouton de Gruyter.

Tywoniw, R., & Crossley, S. A. (2020). Morphological complexity of L2 discourse. In E. Friginal & J. Hardy (Eds). The Routledge Handbook of Corpus Approaches to Discourse Analysis. (pp 269-297). New York: Routledge.

Weischedel, Ralph, Palmer, Martha, Marcus, Mitchell, Hovy, Eduard, Pradhan, Sameer, Ramshaw, Lance, Xue, Nianwen, Taylor, Ann, Kaufman, Jeff, Franchini, Michelle, El-Bachouti, Mohammed, Belvin, Robert, & Houston, Ann. (2013). OntoNotes Release 5.0 [dataset]. Linguistic Data Consortium. https://doi.org/10.35111/XMHB-2B84

White, L. (2003). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: language and cognition, 6*(2), 129-141.

Zaman, A. N. K., Matsakis, P., & Brown, C. (2011). Evaluation of stop word lists in text retrieval using Latent Semantic Indexing. In *2011 Sixth International Conference on Digital Information Management* (pp. 133-136). IEEE.

Zhang, D., & Koda, K. (2012). Contribution of morphological awareness and lexical inferencing ability to L2 vocabulary knowledge and reading comprehension among advanced EFL learners: Testing direct and indirect effects. *Reading and Writing, 25*(5), 1195-1216.

Table 1

Linear model to predict reading ease score

Variable	Relative Importance	Estimate	SE	t	p
(Intercept)		-5.097	0.191	-26.738	< .001
MCI (derivational)	0.203	-0.407	0.031	-13.311	< .001
Word frequency (COCA news logged)	0.200	0.919	0.065	14.121	< .001
Suffix family frequency	0.166	-0.083	0.010	-8.617	< .001
Derivational TTR	0.117	-0.307	0.077	-3.991	< .001
Prefix family frequency	0.103	-0.103	0.022	-4.687	< .001
Root frequency (logged)	0.091	0.534	0.039	13.665	< .001
Prefix frequency	0.083	0.000	0.000	-5.649	< .001
Number of inflections	0.037	1.354	0.159	8.493	< .001

Table 2

Linear model to predict overall ELL proficiency score

Variable	Relative Importance	Estimate	SE	t	p
(Intercept)		0.56664	0.18611	3.045	< .050
Number of morphemes	0.321	0.35747	0.07997	4.47	< .001
Inflectional variety	0.291	0.35534	0.02126	16.71	< .001
Number of roots	0.248	1.87345	0.13909	13.47	< .001
Derivational TTR	0.086	0.24191	0.05096	4.747	< .001
Word frequency (COCA news logged)	0.054	-0.2964	0.05779	-5.129	< .001

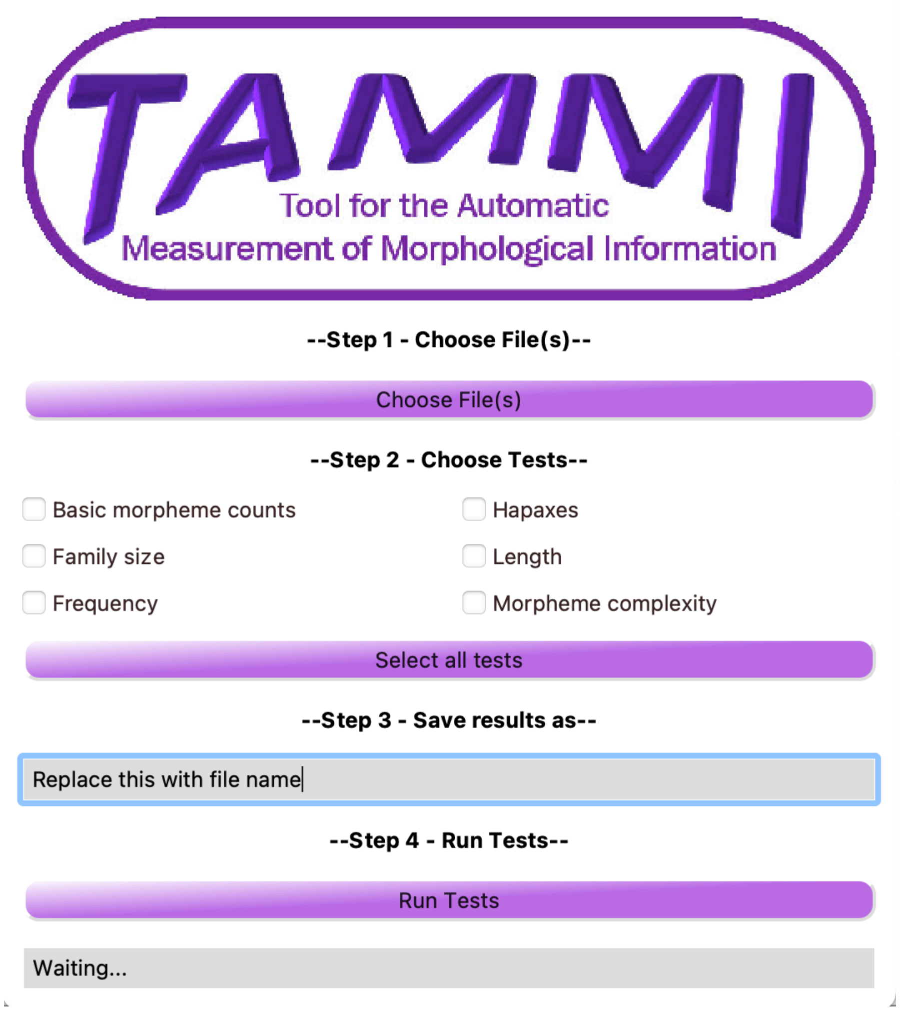

Figure 1. TAMMI Interface

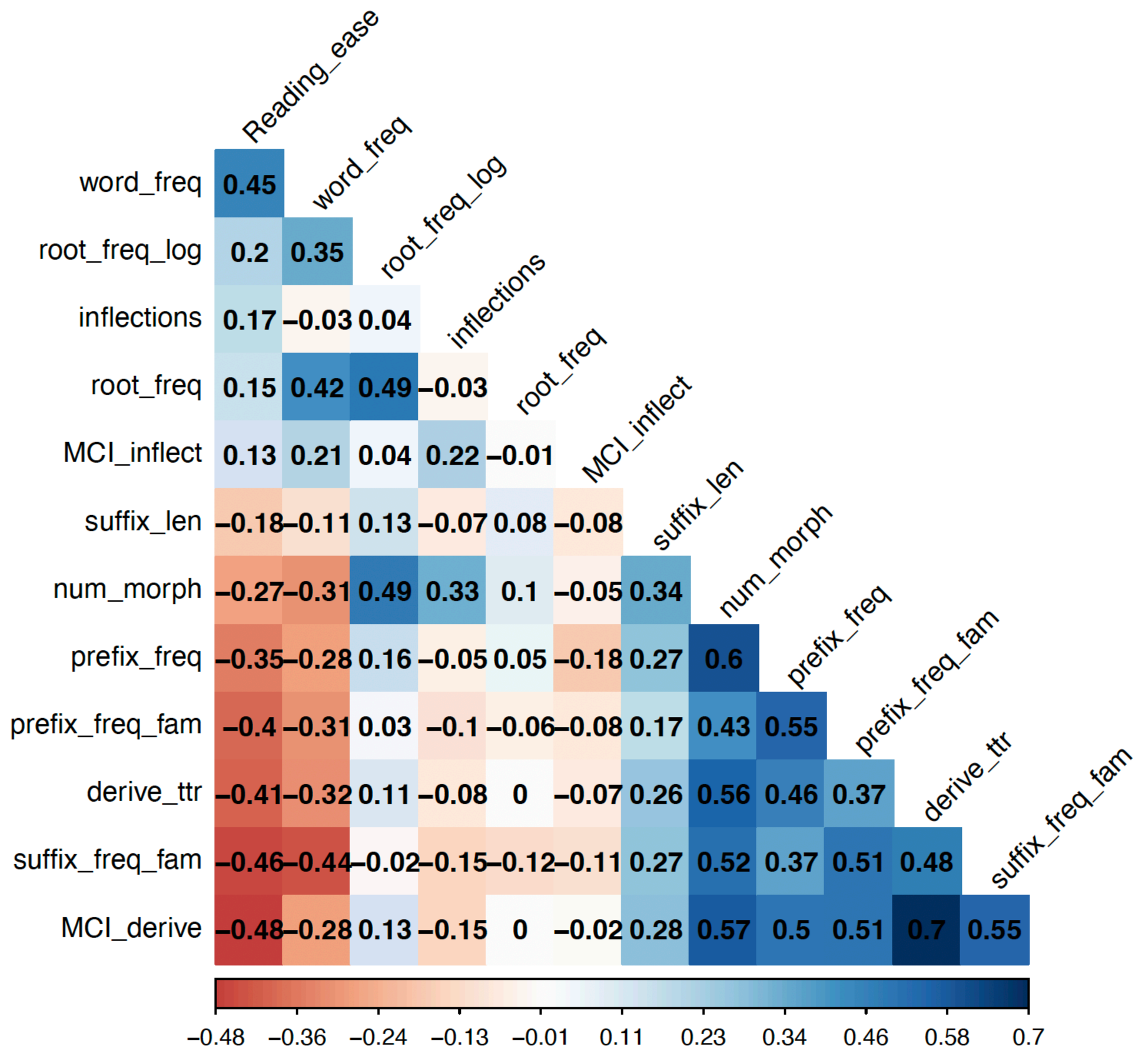

Figure 2: Correlation plot for TAMMI and readability variables used in linear model

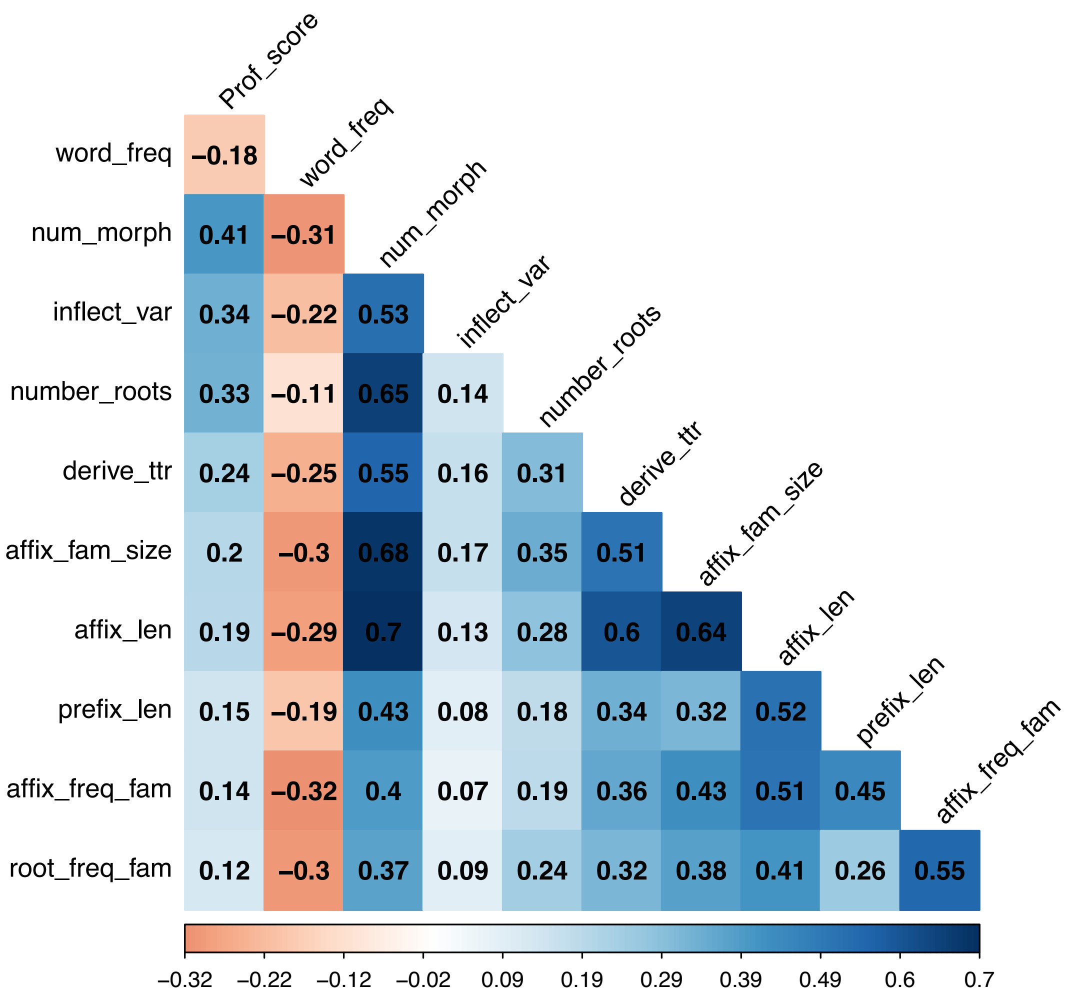

Figure 3: Correlation plot for TAMMI and ELL proficiency variables used in linear model