The
Referential Process

Measures

MEASURES OF THE REFERENTIAL PROCESS

The emotional meaning of written or spoken language may be carried both by the choice of words and the order in which they are used. The Discourse Attributes Analysis Program (DAAP) is designed to capture several dimensions of emotional nuance and meaning as speakers or writers express themselves in ever changing language styles. The program uses a complex procedure to track these language style changes based on the relative usage of many different words, primarily function words, and when they are used.

The three main functions of the referential process (RP) – Arousal, Symbolizing and Reflecting/Reorganizing – are subjective in nature, internal to the speaker, listener, writer or reader, and as such cannot be directly observed or measured. However, as a person speaks or writes, the words they use and the order in which they use them can be examined as indicators of the extent to which each of these underlying functions is active.

There is a distinct linguistic style associated with each of the three referential process (RP) functions. The identification of these styles and their measurement were based originally on judgments of relevant psychological dimensions in sets of texts by trained raters. The dimensions were then modeled using computer assisted procedures based on these judgments. Seemingly paradoxically, we have found that these language styles can be measured by the degree to which the speaker or writer uses various highly frequent words, primarily function words, such as *I, the, and, is, in* rather than by the usage of content words, such as nouns and verbs. Such little words are frequently ignored in many computerized language analysis procedures.

The combined process of judges' ratings and computer modeling has led to the development of three distinct linguistic style variables—each defined by a weighted dictionary—as measures of the three RP functions. The Weighted Arousal List (WRSL) provides a basis for the measure of the Arousal function. The Weighted Referential Activity Dictionary (WRAD) provides a basis for the measure of the Symbolizing function. And the Weighted Reflecting/Reorganizing List (WRRL) provides the basis for our measure of the Reflecting/Reorganizing function. The construction of these and related dictionaries is outlined in the Dictionaries Page. Details are given in the DAAP Technical Manual*.*

The dominance of function words with high frequency allows the measures to be applied with generally high power to verbal material with a wide range of contents. The weighted dictionaries representing the RP functions are relatively small, containing at most 700 words. Many of these words, including the widely used function words mentioned above appear in more than one weighted dictionary, but carry different weights.

Copies of the DAAP program and the RP dictionaries along with installation instructions can be obtained for non-commercial use by contacting installDAAP@daapwrad.org. The distribution also includes unweighted dictionaries representing specific emotions and somatic experiences and other variables.

DaapLAB

A copy of DaapLAB, a standalone application developed for Windows and MacOS, can also be obtained here. DaapLAB implements the fundamental functions of DAAP and enriches it with a filtering algorithm (DaapMAP), allowing the user to automatically highlight the relevant portions of the text. Moreover, DaapLAB© interfaces with Whisper, an open-source AI trained for Speech-to-Text tasks.

The DaapLAB User Guide gives step-by-step instructions on installation and operation.

Download DaapLAB for Windows.

Download DaapLAB for MacOS.

After downloading the appropriate version of DaapLAB for Windows or MacOS, interested users should email daaplab@outlook.com for personal credentials, as indicated in the guide.

The DAAP Smoothing Operator

In general, nearby words in a text have very different dictionary weights so that a word-by-word graph of the dictionary weights in a text is usually incomprehensible. The graph below shows the word-by-word WRAD, WRRL and WRSL dictionary values for Shelley's poem, Ozymandias, with word count on the x-axis and the dictionary weights on the y-axes.
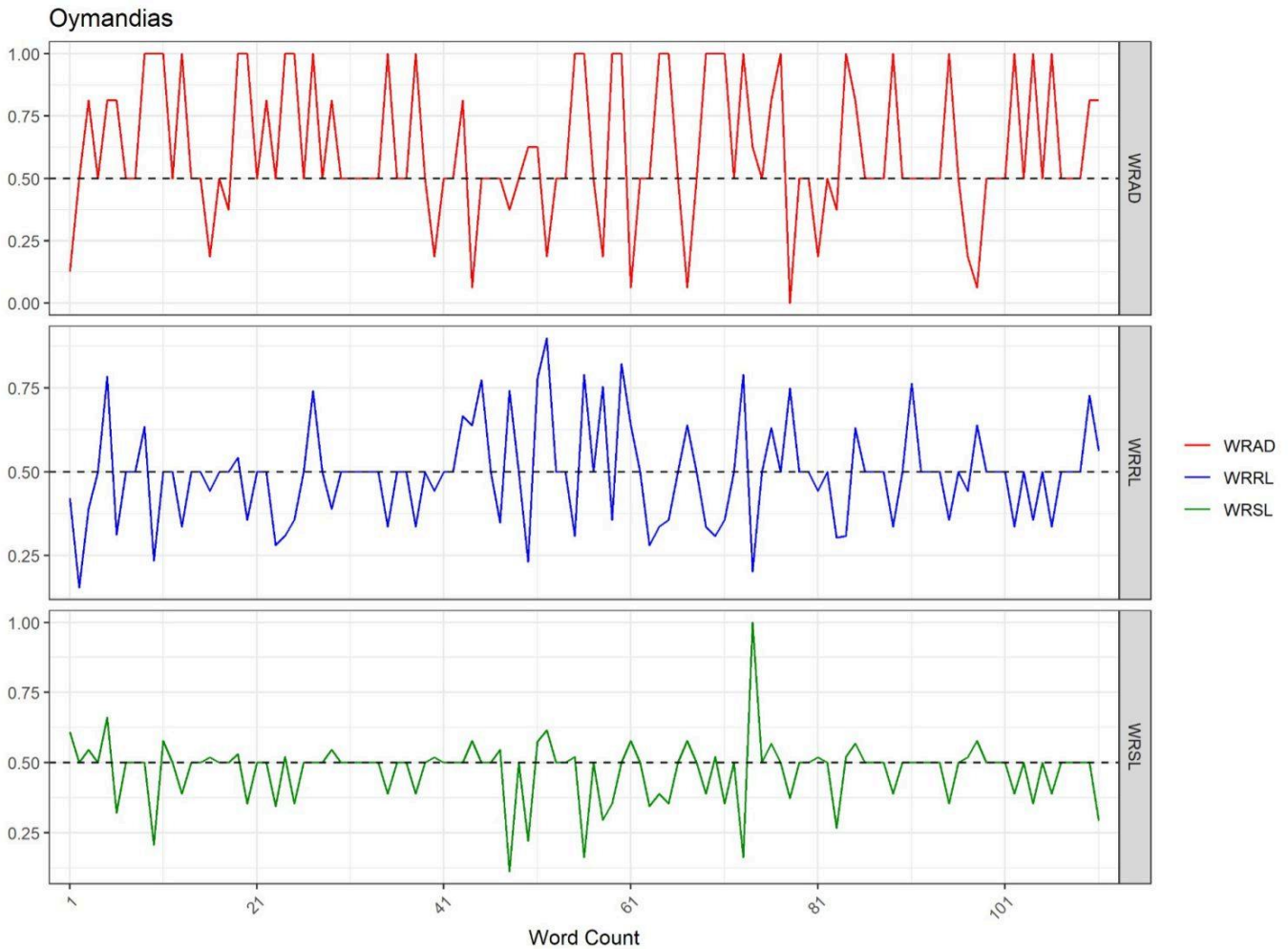
Ozymandias

I met a traveler from an antique land

Who said:[10] Two vast and trunkless legs of stone

Stand in the [20] desert. Near them, on the sand,

Half sunk, a shattered [30] visage lies, whose frown,

And wrinkled lip, and sneer of [40] cold command,

Tell that its sculptor well those passions read [50]

Which yet survive, stamped on these lifeless things,

The hand [60] that mocked them and the heart that fed:

And on [70] the pedestal these words appear:

"My name is Ozymandias, king [80] of kings:

Look on my works, ye Mighty, and despair!" [90]

Nothing beside remains. Round the decay

Of that colossal wreck, [100] boundless and bare

The lone and level sands stretch far [110] away.

The DAAP smoothing operator has been developed to produce smoothly turning functions from such discontinuous data. The graphs of these smoothly turning functions illustrate the ebb and flow of the corresponding linguistic variables, thus providing a more useful graphic representation of a text, while not changing the overall values of the measures. A smoothed version of the Ozymandias text is show in the graphs below, with word count on the x-axis and the smooth WRAD, WRRL and WRSL functions on the y-axis.
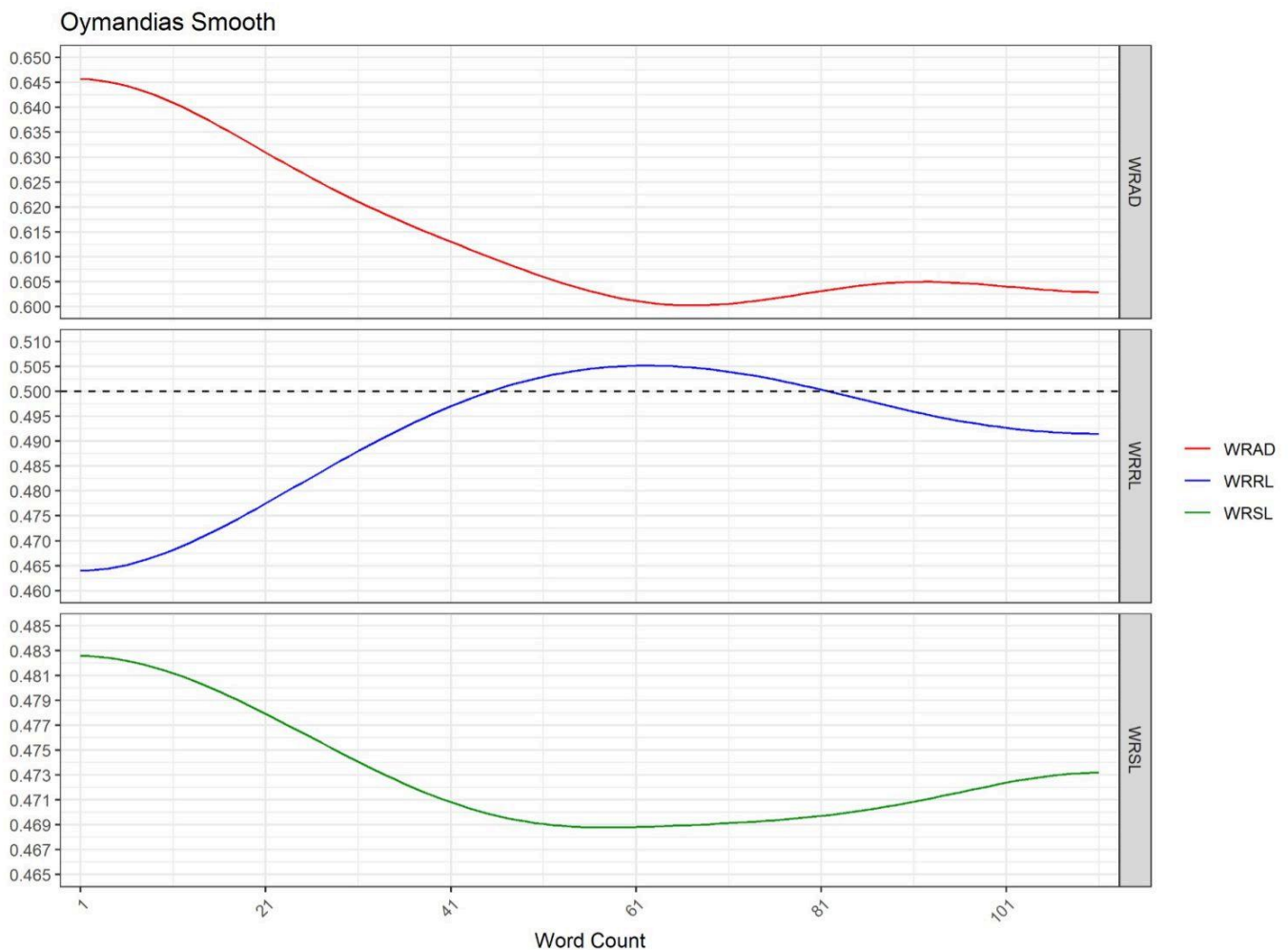
The main ingredient in the smoothing operator is a local weighted averaging procedure, where the local weighted average is based on an exponential function related to the normal curve. We note that the total score and average scores for each smoothed measure are identical to the unsmoothed scores, for each utterance or group of utterances. (Details of the smoothing operator are included in the DAAP Technical Manual. See www.refpro.org for access to the Manual.)

The smoothing operator can also be applied in situations of dialogue or multiple speakers. The next figure illustrates the DAAP smooth graphical image for each of the three RP variables for a psychotherapy session. Patient speech is represented by a relatively thin line and therapist speech is represented by a relatively thick line. These lines are broken at each turn of speech (change of speaker).
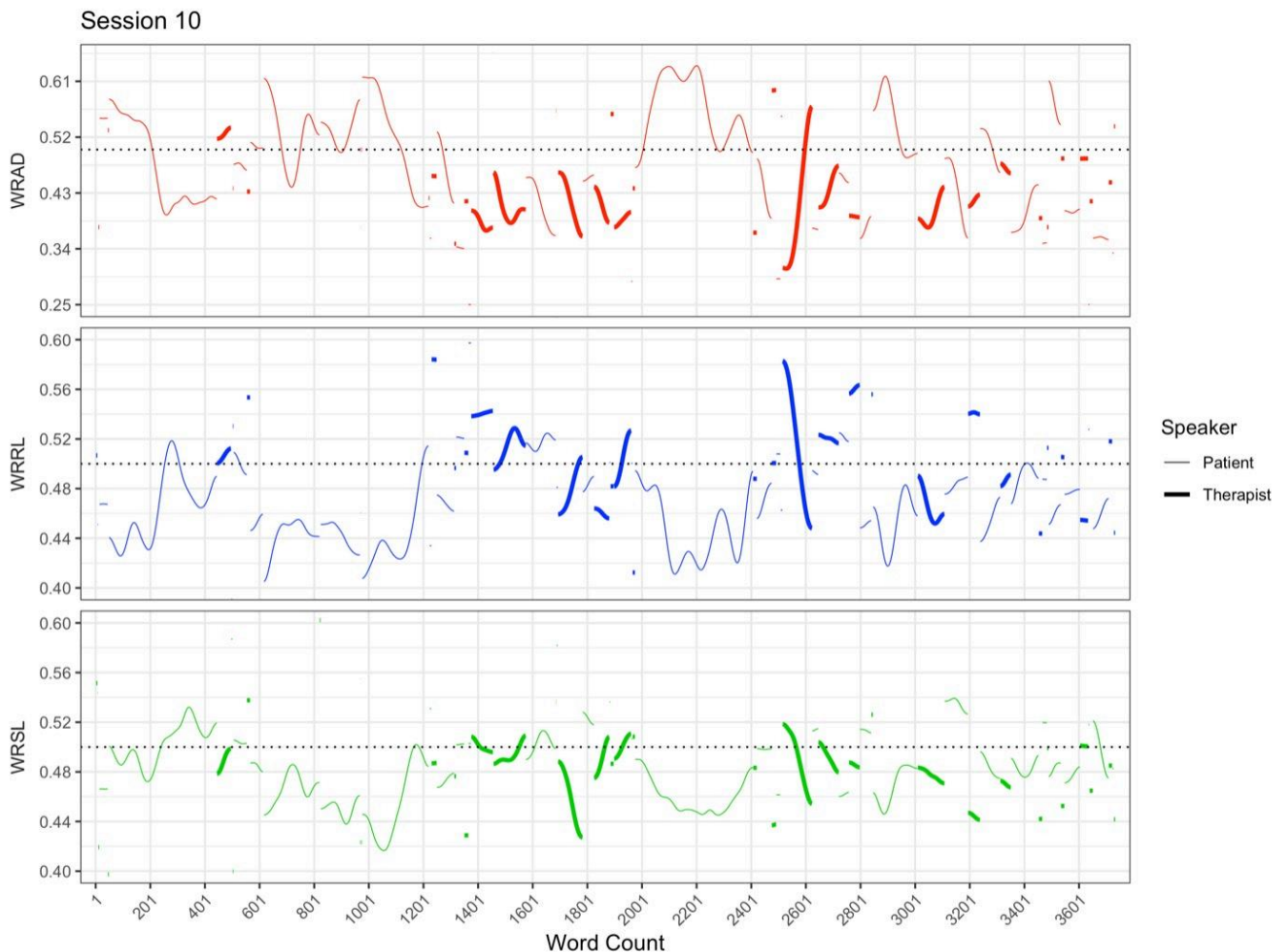
The Neutral Value

For each weighted dictionary, we introduce a number called the Neutral Value, which is based on the scoring procedures, and is set at 0.5. The segment scores were originally scored by judges on a scale of 0 to 10, with a score of 5 considered as *neutral*. At the final step of making each of the dictionaries, the scores were adjusted to range between between 0 and 1.0, thus yielding a neutral value of 0.5 . The neutral value is marked by a horizontal dotted line in each graph.

Word Order and Derived Measures

The emotional significance of a passage depends not only on the words that are spoken or written but also on the order in which they appear. One of the main features of the DAAP system is that it is sensitive to word order. Several new measures that make use of this sensitivity have been developed. They are described on the Derived Measures page.

DERIVED MEASURES AND THE EFFECTS OF WORD ORDER

As described in the first part of the Measures section, the DAAP smoothing operator produces a new measure, the *smooth dictionary value*, for every dictionary, weighted or not, for each word. The power of the smoothing operator to enable readable graphic representations is discussed there. The smoothed values can also be used to produce a set of measures, called the *derived measures,* that go beyond the basic measures of mean dictionary values to provide an additional level of information about the functions of the referential process. The derived measures all use the *neutral value* of 0.5 (See the [Dictionaries](#) page for an explanation of the neutral value). This section covers the major derived measures, as well as the power of the smoothing operator to represent the effects of variation in word order.

Derived Measures

Each of the derived measures produces a number for every type of text by a single speaker, including an utterance, a psychotherapy session or interview as a whole, or any subset or grouping of such units. The same principle applies for written texts such as poems or other literary works.

The derived measures can be used for purposes of comparison between speakers and/or between texts. We emphasize, however, that the computation of the smoothed variables depends on the choice of *smoothing parameters*, so such comparisons are meaningful only if the same values of the smoothing parameters are used in all instances. The smoothing parameters include M (coded as 'param' in DAAP), which determines how many words are included in each iteration of the moving weighted average, and the weighting exponent, Q, (coded as 'paraq' in DAAP), which determines the extent to which the exponential weighting function is flat or pointed). The DAAP program has the smoothing parameters set at the default values of M = 100 and Q = 2.0; users are advised not to change these values.

For each of the weighted dictionaries, WRAD, WRRL and WRSL, DAAP produces the following major derived measures:

1. The *dictionary intensity index* or *mean high dictionary, a* measure of the extent to which the smooth dictionary value for a selected unit of text exceeds the neutral value. This is computed as the mean of the difference between the smoothed scores for words and the neutral value, taken over all words for which the smooth dictionary value is greater than the neutral value of 0.5.
2. The *high dictionary proportion,* the proportion of words in the selected text for which the smooth dictionary value is greater than the neutral value of 0.5.

For example, for the WRAD, the Mean High WRAD (MHWRAD) is the *mean of the difference* between the smoothed dictionary values of WRAD and the neutral value, taken over all the words for which the smoothed WRAD is greater than the neutral value. We can understand this variable as representing how high the smoothed WRAD is when it is high. The High WRAD Proportion (HPWRAD) is the *proportion of words* in the text for which the smoothed dictionary value of WRAD is greater than the neutral value of 0.5. We can understand this

variable as representing the proportion of an utterance, or set of utterances, for which the smoothed WRAD function is high. The same principles apply for the other weighted dictionaries.

In metaphoric terms we can understand these variables in terms of the speed of persons who cover a certain distance sometimes running and sometimes walking. The Mean High Running measure (*MHRunning)* would tell us **how fast** the person is going when they are running; the High Running Proportion (HPRunning) would tell us **what proportion of the distance the person is covering** when they are running**.**

The covariations. For each pair of dictionaries, DAAP produces a *covariation* which is a measure of *the extent to which two dictionary values are simultaneously high and simultaneously low*. The covariation can in principle be computed for any pair of dictionaries, weighted or unweighted. Here, as in the computation of the correlation coefficient, a measure is considered to be high if it is greater than the measure *mean*, where this mean is computed as the average measure for the text under consideration.

The covariations provide an additional level of information concerning the referential process beyond that provided by the measures themselves. For example, a high *negative* covariation of WRAD and WRRL would indicate that a speaker is separately immersed in describing an experience and reflecting on it. There is both a theoretical basis and preliminary evidence to suggest that a high negative WRAD/WRRL covariation in patient language is associated with effective psychotherapy.

THE EFFECTS OF WORD ORDER

One consequence of the smoothing operator is that the word-by-word smoothed dictionary values are sensitive to the order in which the words are spoken or written. The smoothed score for any individual word depends on the dictionary scores of the nearby surrounding words and the order in which they appear. To illustrate these effects, we present two segments of spoken language; one shown as spoken in a psychotherapy session; the other containing the exact same words in somewhat different order. The numbers in square brackets show word count corresponding to the tick marks on the x-axis of the corresponding graphs.

(Segment A; PCX1530; Turn 13)] Mm, that just makes me think again of the feeling I find very hard to describe, that was certainly [20] stronger in the few weeks right after I came out of the hospital, than it is now. but when [40] I was nursing her and looking at her, and almost looking at her, looking at her, thinking, I've got [60] to be able to see her, and feeling I just couldn't. and I suppose that was partly because I [80] thought by looking at her, I'd feel a warm feeling toward her, and I just really couldn't.
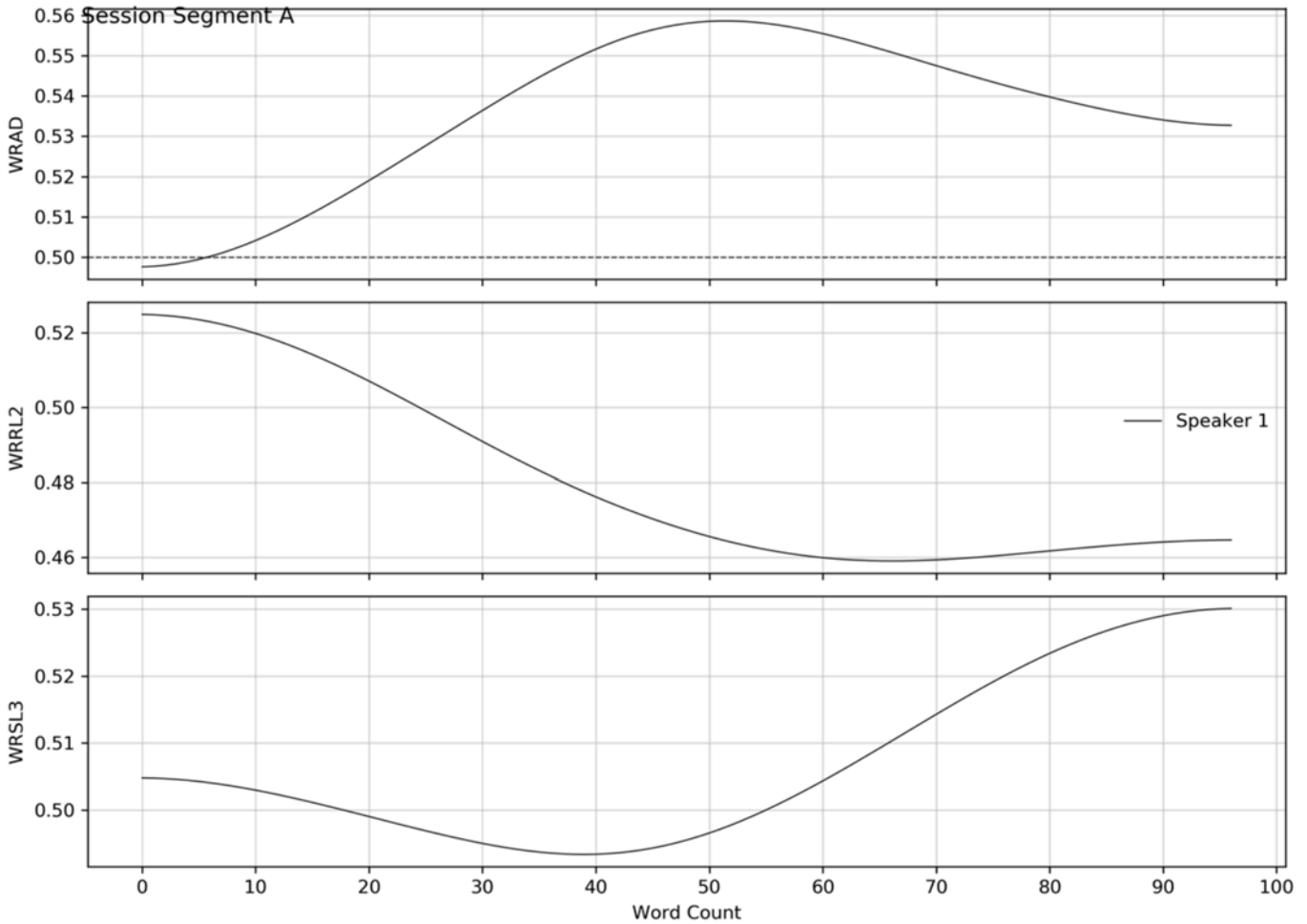
The second segment uses the same words, but in somewhat different order.

(Segment B) That feeling was certainly stronger in the few weeks right after I came out of the hospital, than it is [20] now and I suppose that was partly because I thought by looking at her, I'd feel a warm feeling [40] toward her, but I just really couldn't. and mm I find the feeling very hard to describe I just [60] couldn't and.... that just makes me think again of when I was nursing her and looking at her and [80] almost looking at her, looking at her, thinking, I've got to be able to see her,

While the same words are used in the second segment the quality of the communication is quite different as indicated by the smoothed graphs.
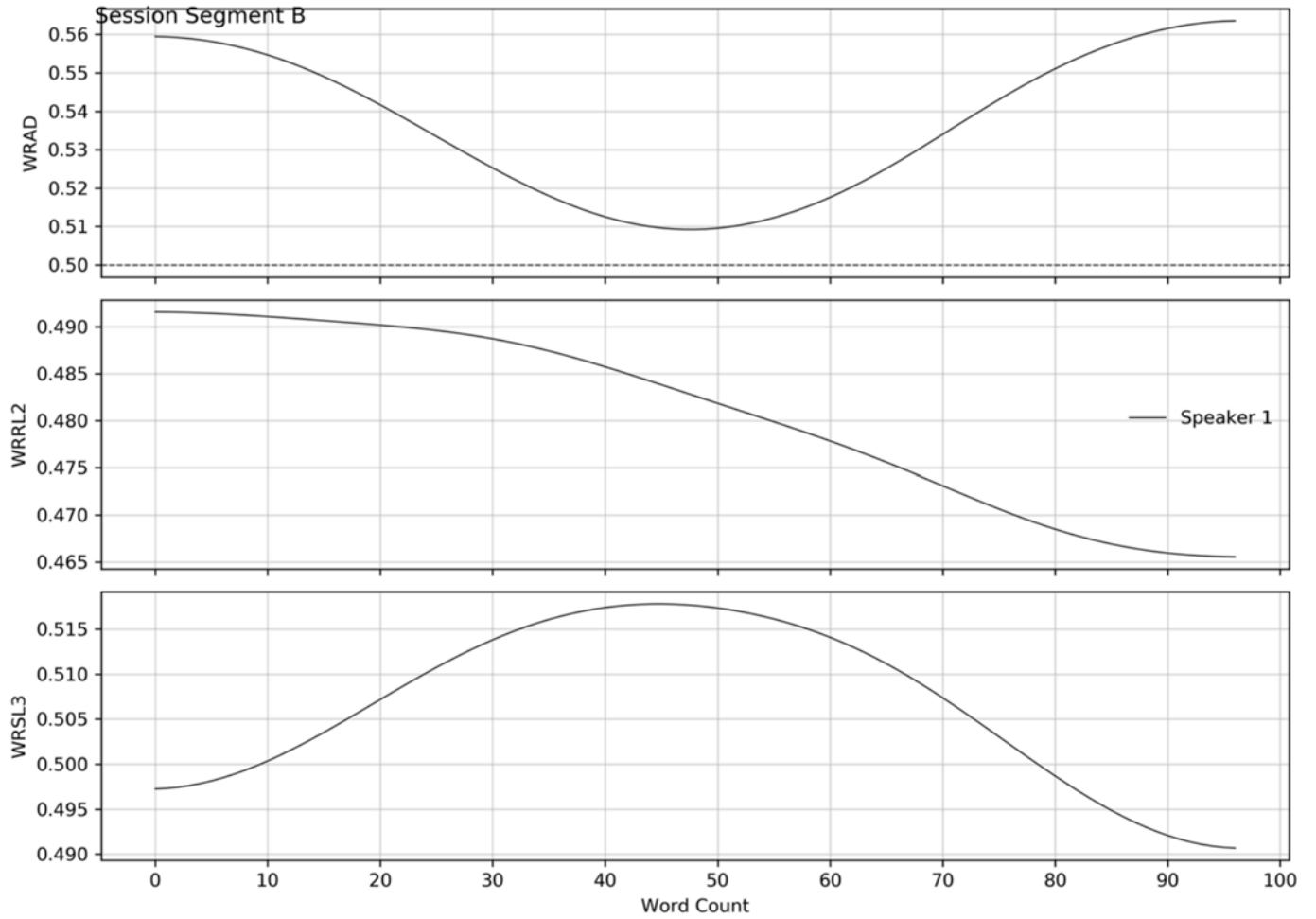
There are also quantitative overall differences that are seen for the derived measures, which are marked in Table 1 by an asterisk. These differences are not seen for the standard measures of Word Count or Mean Dictionary score.

Measure	Segment A	Segment B
Number of Words	97	97
Mean WRAD	0.536	0.536
Mean High WRAD*	0.039	0.036
High WRAD Proportion*	0.938	1.0
Mean WRRL	0.481	0.481
Mean High WRRL*	0.016	0.0
High WRRL Proportion*	0.258	0.0
Mean WRSL	0.507	0.507
Mean High WRSL*	0.013	0.011
High WRSL Proportion*	0.619	0.711
WRAD_WRRL Covar*	-0.883	-0.242
WRAD_WRSL Covar*	-0.426	-0.423
WRRL_WRSL Covar*	-0.350	0.423

The
Referential Process

Home   Theory   Measures   Events and Publications

People