**Overview of the Textual Paralanguage (PARA) Classifier**

This software is designed to identify nonverbal parts of speech expressed in text-based communication. Textual paralanguage (TPL) is defined as written manifestations of nonverbal audible, tactile, and visual elements that supplement or replace written language and that can be expressed through words, symbols, images, punctuation, demarcations, or any combination of these elements (see Luangrath et al. 2017). Textual paralanguage is categorized into five main categories of voice qualities (VQ), vocalizations (VS), tactile kinesics (TK), visual kinesics (VK), and artifacts (A). While most text analytic tools rely on nuances in the meaning of actual words themselves, this tool identifies the extratextual features of written communication that communicate nonverbal expression.

The TPL classifier relies on a panel of internal dictionaries that define which words/symbols/images should be counted in the target text files. The TPL classifier is designed to accept written or transcribed verbal text which has been stored as a digital, machine- readable file in standard .csv format and is compatible with PC or Mac computers. Files should be uploaded with UTF8 encoding for proper display of images and emojis. The software can process text on a line by line basis within columns in spreadsheets. During operation, the classifier engages in iterative processing by systematically expanding or contracting word forms to recheck against the dictionaries to determine whether words or characters are indications of nonverbal expressions in text. As the text file is being processed, counts for various structural composition elements are identified.

For each text file, 22 output variables are written as one line of data to an output file. This data record includes the TPL\_Total (a total count of all TPL elements), Pitch, Rhythm, Stress, Emphasis, Tempo, Volume, Censorship, Spelling, Alternants, Differentiators, Tactile\_Emojis, Alphahaptics, Tactile\_Emoticons, Bodily\_Emoticons, Bodily\_Emojis, Alphakinesics,

Nonbodily\_Emojis, Formatting, Nonbodily\_Emoticons; see Table 1 and Table 2). Combining certain elements, the software also produces a Emoji\_Total(summation of Tactile\_Emojis, Bodily\_Emojis, and Nonbodily\_Emojis) and a

Emoticon\_Total (summation of Tactile\_Emoticons, Bodily\_Emoticons, and Nonbodily\_Emoticons). A complete list of the standard output variables is included in Table 2.

*Table 1. Textual Paralinguistic Category Definitions (Luangrath, Peck, and Barger 2017)*

|  |  |
| --- | --- |
| **Voice Qualities** | Characteristics of the sound of the words being communicated that have to do with how the words should be spoken |
| **Vocalizations** | Utterances, fillers, terms, or sounds that can be spoken or produced by the body that result in an audible noise that is recognizable (not necessarily a “word”) |
| **Tactile Kinesics** | Nonverbal communication related to physical, haptic interaction with others |
| **Visual Kinesics** | Nonverbal communication related to movement of any part of the body or the body as a whole |
| **Artifacts** | The presentational, formatting, and stylistic elements of a message |

*Table 2. A complete list of the standard output variables (based on Luangrath, Peck, and Barger 2017)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Superordinate**  **Level Sensory**  **Features** | Auditory | | Tactile | Visual | | **Aggregate Variables** |
| **Textual**  **Paralinguistic**  **Categories** | **Voice Qualities** | **Vocalizations** | **Tactile Kinesics** | **Visual Kinesics** | **Artifacts** |
| **Subordinate**  **Level**  **Nonverbal**  **Features** | *Pitch*  *Rhythm*  *Stress*  *Emphasis*  *Tempo*  *Volume*  *Censorship*  *Spelling* | *Alternants Differentiators* | *Tactile\_Emojis*  *Tactile\_Emoticons*  *Alphahaptics* | *Bodily\_Emojis*  *Bodily\_Emoticons*  *Alphakinesics* | *Nonbodily\_Emojis*  *Nonbodily\_Emoticons Formatting* | *Emoji\_Total*  *Emoticon\_Total*  *TPL\_Total* |

*Note: Italic and underlined elements are the output variables produced by the software.*