

# 1 List of all features

Table 1: List of all features with their description and category

| Feature                | Description  | Category                  |
|------------------------|--|---------------------------|
| amount_adj_sli         | Number of adjectives in the slides                                       | Syntactic<br>(word types) |
| avg_adj_sli            | Average number of adjectives per line in the slides                      | Syntactic<br>(word types) |
| ratio_adj_sli          | Relation of the number of adjectives to unfiltered tokens in the slides  | Syntactic<br>(word types) |
| amount_adpos_sli       | Number of adpositions in the slides                                      | Syntactic<br>(word types) |
| avg_adpos_sli          | Average number of adpositions per line in the slides                     | Syntactic<br>(word types) |
| ratio_adpos_sli        | Relation of the number of adpositions to unfiltered tokens in the slides | Syntactic<br>(word types) |
| amount_noun_sli        | Number of nouns in the slides  | Syntactic<br>(word types) |
| avg_noun_sli           | Average number of nouns per line in the slides                           | Syntactic<br>(word types) |
| ratio_noun_sli         | Relation of the number of nouns to unfiltered tokens in the slides       | Syntactic<br>(word types) |
| amount_pronoun_sli     | Number of pronouns in the slides   | Syntactic<br>(word types) |
| avg_pronoun_sli        | Average number of pronouns per line in the slides                        | Syntactic<br>(word types) |
| ratio_pronoun_sli      | Relation of the number of pronouns to unfiltered tokens in the slides    | Syntactic<br>(word types) |
| ratio_pronoun_noun_sli | Relation of the number of pronouns to nouns in the slides                | Syntactic<br>(word types) |
| amount_verb_sli        | Number of verbs in the slides  | Syntactic<br>(word types) |
| avg_verb_sli           | Average number of verbs per line in the slides                           | Syntactic<br>(word types) |
| ratio_verb_sli         | Relation of the number of verbs to unfiltered tokens in the slides       | Syntactic<br>(word types) |
| amount_main_verb_sli   | Number of main verbs in the slides                                       | Syntactic<br>(word types) |
| avg_main_verb_sli      | Average number of main verbs per line in the slides                      | Syntactic<br>(word types) |
| ratio_main_verb_sli    | Relation of the number of main verbs to unfiltered tokens in the slides  | Syntactic<br>(word types) |
| amount_aux_sli         | Number of auxiliaries in the slides                                      | Syntactic<br>(word types) |
| avg_aux_sli            | Average number of auxiliaries per line in the slides                     | Syntactic<br>(word types) |
| ratio_aux_sli          | Relation of the number of auxiliaries to unfiltered tokens in the slides | Syntactic<br>(word types) |
| amount_adverb_sli      | Number of adverbs in the slides  | Syntactic<br>(word types) |
| avg_adverb_sli         | Average number of adverbs per line in the slides                         | Syntactic<br>(word types) |
| ratio_adverb_sli       | Relation of the number of adverbs to unfiltered tokens in the slides     | Syntactic<br>(word types) |

Continuation on the next page

Table 1 – Continuation

| Feature                  | Description  | Category                  |
|--------------------------|--|---------------------------|
| amount_coord_conj_sli    | Number of coordinate conjunctions in the slides                                      | Syntactic<br>(word types) |
| avg_coord_conj_sli       | Average number of coordinate conjunctions per line in the slides                     | Syntactic<br>(word types) |
| ratio_coord_conj_sli     | Relation of the number of coordinate conjunctions to unfiltered tokens in the slides | Syntactic<br>(word types) |
| amount_determiner_sli    | Number of determiner in the slides   | Syntactic<br>(word types) |
| avg_determiner_sli       | Average number of determiner per line in the slides                                  | Syntactic<br>(word types) |
| ratio_determiner_sli     | Relation of the number of determiner to unfiltered tokens in the slides              | Syntactic<br>(word types) |
| amount_interj_sli        | Number of interjections in the slides  | Syntactic<br>(word types) |
| avg_interj_sli           | Average number of interjections per line in the slides                               | Syntactic<br>(word types) |
| ratio_interj_sli         | Relation of the number of interjections to unfiltered tokens in the slides           | Syntactic<br>(word types) |
| amount_num_sli           | Number of numbers in the slides  | Syntactic<br>(word types) |
| avg_num_sli              | Average number of numbers per line in the slides                                     | Syntactic<br>(word types) |
| ratio_num_sli            | Relation of the number of numbers to unfiltered tokens in the slides                 | Syntactic<br>(word types) |
| amount_particle_sli      | Number of particles in the slides  | Syntactic<br>(word types) |
| avg_particle_sli         | Average number of particles per line in the slides                                   | Syntactic<br>(word types) |
| ratio_particle_sli       | Relation of the number of particles to unfiltered tokens in the slides               | Syntactic<br>(word types) |
| amount_subord_conj_sli   | Number of particles in the slides  | Syntactic<br>(word types) |
| avg_subord_conj_sli      | Average number of particles per line in the slides                                   | Syntactic<br>(word types) |
| ratio_subord_conj_sli    | Relation of the number of particles to unfiltered tokens in the slides               | Syntactic<br>(word types) |
| amount_foreign_sli       | Number of foreign words in the slides  | Syntactic<br>(word types) |
| avg_foreign_sli          | Average number of foreign words per line in the slides                               | Syntactic<br>(word types) |
| ratio_foreign_sli        | Relation of the number of foreign words to unfiltered tokens in the slides           | Syntactic<br>(word types) |
| amount_content_word_sli  | Number of content words in the slides  | Syntactic<br>(word types) |
| avg_content_word_sli     | Average number of content words per line in the slides                               | Syntactic<br>(word types) |
| ratio_content_word_sli   | Relation of the number of content words to unfiltered tokens in the slides           | Syntactic<br>(word types) |
| amount_function_word_sli | Number of function words in the slides   | Syntactic<br>(word types) |
| avg_function_word_sli    | Average number of function words per line in the slides                              | Syntactic<br>(word types) |
| ratio_function_word_sli  | Relation of the number of function words to unfiltered tokens in the slides          | Syntactic<br>(word types) |

Continuation on the next page

Table 1 – Continuation

| Feature                | Description   | Category                  |
|------------------------|---|---------------------------|
| amount_filtered_sli    | Number of filtered words in the slides                                      | Syntactic<br>(word types) |
| avg_filtered_sli       | Average number of filtered words per line in the slides                     | Syntactic<br>(word types) |
| ratio_filtered_sli     | Relation of the number of filtered words to unfiltered tokens in the slides | Syntactic<br>(word types) |
| amount_adj_tra         | Number of adjectives in the srt   | Syntactic<br>(word types) |
| avg_adj_tra            | Average number of adjectives per sentence in the srt                        | Syntactic<br>(word types) |
| ratio_adj_tra          | Relation of the number of adjectives to unfiltered tokens in the srt        | Syntactic<br>(word types) |
| amount_adpos_tra       | Number of adpositions in the srt  | Syntactic<br>(word types) |
| avg_adpos_tra          | Average number of adpositions per sentence in the srt                       | Syntactic<br>(word types) |
| ratio_adpos_tra        | Relation of the number of adpositions to unfiltered tokens in the srt       | Syntactic<br>(word types) |
| amount_noun_tra        | Number of nouns in the srt  | Syntactic<br>(word types) |
| avg_noun_tra           | Average number of nouns per sentence in the srt                             | Syntactic<br>(word types) |
| ratio_noun_tra         | Relation of the number of nouns to unfiltered tokens in the srt             | Syntactic<br>(word types) |
| amount_pronoun_tra     | Number of pronouns in the srt   | Syntactic<br>(word types) |
| avg_pronoun_tra        | Average number of pronouns per sentence in the srt                          | Syntactic<br>(word types) |
| ratio_pronoun_tra      | Relation of the number of pronouns to unfiltered tokens in the srt          | Syntactic<br>(word types) |
| ratio_pronoun_noun_tra | Relation of the number of pronouns to nouns in the srt                      | Syntactic<br>(word types) |
| amount_verb_tra        | Number of verbs in the srt  | Syntactic<br>(word types) |
| avg_verb_tra           | Average number of verbs per sentence in the srt                             | Syntactic<br>(word types) |
| ratio_verb_tra         | Relation of the number of verbs to unfiltered tokens in the srt             | Syntactic<br>(word types) |
| amount_main_verb_tra   | Number of main verbs in the srt   | Syntactic<br>(word types) |
| avg_main_verb_tra      | Average number of main verbs per sentence in the srt                        | Syntactic<br>(word types) |
| ratio_main_verb_tra    | Relation of the number of main verbs to unfiltered tokens in the srt        | Syntactic<br>(word types) |
| amount_aux_tra         | Number of auxiliaries in the srt  | Syntactic<br>(word types) |
| avg_aux_tra            | Average number of auxiliaries per sentence in the srt                       | Syntactic<br>(word types) |
| ratio_aux_tra          | Relation of the number of auxiliaries to unfiltered tokens in the srt       | Syntactic<br>(word types) |
| amount_adverb_tra      | Number of adverbs in the srt  | Syntactic<br>(word types) |
| avg_adverb_tra         | Average number of adverbs per sentence in the srt                           | Syntactic<br>(word types) |

Continuation on the next page

Table 1 – Continuation

| Feature                  | Description   | Category                  |
|--------------------------|---|---------------------------|
| ratio_adverb_tra         | Relation of the number of adverbs to unfiltered tokens in the srt                 | Syntactic<br>(word types) |
| amount_coord_conj_tra    | Number of coordinate conjunctions in the srt                                      | Syntactic<br>(word types) |
| avg_coord_conj_tra       | Average number of coordinate conjunctions per sentence in the srt                 | Syntactic<br>(word types) |
| ratio_coord_conj_tra     | Relation of the number of coordinate conjunctions to unfiltered tokens in the srt | Syntactic<br>(word types) |
| amount_determiner_tra    | Number of determiner in the srt   | Syntactic<br>(word types) |
| avg_determiner_tra       | Average number of determiner per sentence in the srt                              | Syntactic<br>(word types) |
| ratio_determiner_tra     | Relation of the number of determiner to unfiltered tokens in the srt              | Syntactic<br>(word types) |
| amount_interj_tra        | Number of interjections in the srt  | Syntactic<br>(word types) |
| avg_interj_tra           | Average number of interjections per sentence in the srt                           | Syntactic<br>(word types) |
| ratio_interj_tra         | Relation of the number of interjections to unfiltered tokens in the srt           | Syntactic<br>(word types) |
| amount_num_tra           | Number of numbers in the srt  | Syntactic<br>(word types) |
| avg_num_tra              | Average number of numbers per sentence in the srt                                 | Syntactic<br>(word types) |
| ratio_num_tra            | Relation of the number of numbers to unfiltered tokens in the srt                 | Syntactic<br>(word types) |
| amount_particle_tra      | Number of particles in the srt  | Syntactic<br>(word types) |
| avg_particle_tra         | Average number of particles per sentence in the srt                               | Syntactic<br>(word types) |
| ratio_particle_tra       | Relation of the number of particles to unfiltered tokens in the srt               | Syntactic<br>(word types) |
| amount_subord_conj_tra   | Number of particles in the srt  | Syntactic<br>(word types) |
| avg_subord_conj_tra      | Average number of particles per sentence in the srt                               | Syntactic<br>(word types) |
| ratio_subord_conj_tra    | Relation of the number of particles to unfiltered tokens in the srt               | Syntactic<br>(word types) |
| amount_foreign_tra       | Number of foreign words in the srt  | Syntactic<br>(word types) |
| avg_foreign_tra          | Average number of foreign words per sentence in the srt                           | Syntactic<br>(word types) |
| ratio_foreign_tra        | Relation of the number of foreign words to unfiltered tokens in the srt           | Syntactic<br>(word types) |
| amount_content_word_tra  | Number of content words in the srt  | Syntactic<br>(word types) |
| avg_content_word_tra     | Average number of content words per sentence in the srt                           | Syntactic<br>(word types) |
| ratio_content_word_tra   | Relation of the number of content words to unfiltered tokens in the srt           | Syntactic<br>(word types) |
| amount_function_word_tra | Number of function words in the srt   | Syntactic<br>(word types) |
| avg_function_word_tra    | Average number of function words per sentence in the srt                          | Syntactic<br>(word types) |

Continuation on the next page

Table 1 – Continuation

| Feature                  | Description   | Category               |
|--------------------------|---|------------------------|
| ratio_function_word_tra  | Relation of the number of function words to unfiltered tokens in the srt              | Syntactic (word types) |
| amount_filtered_tra      | Number of filtered words in the srt   | Syntactic (word types) |
| avg_filtered_tra         | Average number of filtered words per sentence in the srt                              | Syntactic (word types) |
| ratio_filtered_tra       | Relation of the number of filtered words to unfiltered tokens in the srt              | Syntactic (word types) |
| sim_pres_sli             | Number of Simple Present tenses in the slides   | Syntactic (temporal)   |
| ratio_sim_pres_sli       | Relation of number of Simple Present tenses to all phrases in the slides              | Syntactic (temporal)   |
| pres_prog_sli            | Number of Present Progressive tenses in the slides                                    | Syntactic (temporal)   |
| ratio_pres_prog_sli      | Relation of number of Present Progressive tenses to all phrases in the slides         | Syntactic (temporal)   |
| pres_perf_sli            | Number of Present Perfect tenses in the slides  | Syntactic (temporal)   |
| ratio_pres_perf_sli      | Relation of number of Present Perfect tenses to all phrases in the slides             | Syntactic (temporal)   |
| pres_perf_prog_sli       | Number of Present Perfect Progressive tenses in the slides                            | Syntactic (temporal)   |
| ratio_pres_perf_prog_sli | Relation of number of Present Perfect Progressive tenses to all phrases in the slides | Syntactic (temporal)   |
| sim_pas_sli              | Number of Simple Past tenses in the slides  | Syntactic (temporal)   |
| ratio_sim_pas_sli        | Relation of number of Simple Past tenses to all phrases in the slides                 | Syntactic (temporal)   |
| pas_prog_sli             | Number of Past Progressive tenses in the slides                                       | Syntactic (temporal)   |
| ratio_pas_prog_sli       | Relation of number of Past Progressive tenses to all phrases in the slides            | Syntactic (temporal)   |
| pas_perf_sli             | Number of Past Perfect tenses in the slides   | Syntactic (temporal)   |
| ratio_pas_perf_sli       | Relation of number of Past Perfect tenses to all phrases in the slides                | Syntactic (temporal)   |
| pas_perf_prog_sli        | Number of Past Perfect Progressive tenses in the slides                               | Syntactic (temporal)   |
| ratio_pas_perf_prog_sli  | Relation of number of Past Perfect Progressive tenses to all phrases in the slides    | Syntactic (temporal)   |
| will_sli                 | Number of Will-Future tenses in the slides  | Syntactic (temporal)   |
| ratio_will_sli           | Relation of number of Will-Future tenses to all phrases in the slides                 | Syntactic (temporal)   |
| fu_prog_sli              | Number of Future Progressive tenses in the slides                                     | Syntactic (temporal)   |
| ratio_fu_prog_sli        | Relation of number of Future Progressive tenses to all phrases in the slides          | Syntactic (temporal)   |
| fu_perf_sli              | Number of Future Perfect tenses in the slides   | Syntactic (temporal)   |
| ratio_fu_perf_sli        | Relation of number of Future Perfect tenses to all phrases in the slides              | Syntactic (temporal)   |
| fu_perf_prog_sli         | Number of Future Perfect Progressive tenses in the slides                             | Syntactic (temporal)   |

Continuation on the next page

Table 1 – Continuation

| Feature                  | Description   | Category             |
|--------------------------|---|----------------------|
| ratio_fu_perf_prog_sli   | Relation of number of Future Perfect Progressive tenses to all phrases in the slides      | Syntactic (temporal) |
| cond_sim_sli             | Number of Conditional Simple tenses in the slides   | Syntactic (temporal) |
| ratio_cond_sim_sli       | Relation of number of Conditional Simple tenses to all phrases in the slides              | Syntactic (temporal) |
| cond_prog_sli            | Number of Conditional Progressive tenses in the slides                                    | Syntactic (temporal) |
| ratio_cond_prog_sli      | Relation of number of Conditional Progressive tenses to all phrases in the slides         | Syntactic (temporal) |
| cond_perf_sli            | Number of Conditional Perfect tenses in the slides  | Syntactic (temporal) |
| ratio_cond_perf_sli      | Relation of number of Conditional Perfect tenses to all phrases in the slides             | Syntactic (temporal) |
| cond_perf_prog_sli       | Number of Conditional Perfect Progressive tenses in the slides                            | Syntactic (temporal) |
| ratio_cond_perf_prog_sli | Relation of number of Conditional Perfect Progressive tenses to all phrases in the slides | Syntactic (temporal) |
| gerund_sli               | Number of Gerund/Present Participle tenses in the slides                                  | Syntactic (temporal) |
| ratio_gerund_sli         | Relation of number of Gerund/Present Participle tenses to all phrases in the slides       | Syntactic (temporal) |
| perf_part_sli            | Number of Perfect Participle tenses in the slides   | Syntactic (temporal) |
| ratio_perf_part_sli      | Relation of number of Perfect Participle tenses to all phrases in the slides              | Syntactic (temporal) |
| inf_sli                  | Number of Present Infinitive tenses in the slides   | Syntactic (temporal) |
| ratio_inf_sli            | Relation of number of Present Infinitive tenses to all phrases in the slides              | Syntactic (temporal) |
| perf_inf_sli             | Number of Perfect Infinitive tenses in the slides   | Syntactic (temporal) |
| ratio_perf_inf_sli       | Relation of number of Perfect Infinitive tenses to all phrases in the slides              | Syntactic (temporal) |
| active_sli               | Number of active verb forms in the slides   | Syntactic (temporal) |
| ratio_active_sli         | Relation of number of active verb forms to all phrases in the slides                      | Syntactic (temporal) |
| passive_sli              | Number of passive verb forms in the slides  | Syntactic (temporal) |
| ratio_passive_sli        | Relation of number of passive verb forms to all phrases in the slides                     | Syntactic (temporal) |
| sim_pres_tra             | Number of Simple Present tenses in the srt  | Syntactic (temporal) |
| ratio_sim_pres_tra       | Relation of number of Simple Present tenses to all phrases in the srt                     | Syntactic (temporal) |
| pres_prog_tra            | Number of Present Progressive tenses in the srt   | Syntactic (temporal) |
| ratio_pres_prog_tra      | Relation of number of Present Progressive tenses to all phrases in the srt                | Syntactic (temporal) |
| pres_perf_tra            | Number of Present Perfect tenses in the srt   | Syntactic (temporal) |
| ratio_pres_perf_tra      | Relation of number of Present Perfect tenses to all phrases in the srt                    | Syntactic (temporal) |

Continuation on the next page

Table 1 – Continuation

| Feature                  | Description  | Category             |
|--------------------------|--|----------------------|
| pres_perf_prog_tra       | Number of Present Perfect Progressive tenses in the srt                                | Syntactic (temporal) |
| ratio_pres_perf_prog_tra | Relation of number of Present Perfect Progressive tenses to all phrases in the srt     | Syntactic (temporal) |
| sim_pas_tra              | Number of Simple Past tenses in the srt  | Syntactic (temporal) |
| ratio_sim_pas_tra        | Relation of number of Simple Past tenses to all phrases in the srt                     | Syntactic (temporal) |
| pas_prog_tra             | Number of Past Progressive tenses in the srt   | Syntactic (temporal) |
| ratio_pas_prog_tra       | Relation of number of Past Progressive tenses to all phrases in the srt                | Syntactic (temporal) |
| pas_perf_tra             | Number of Past Perfect tenses in the srt   | Syntactic (temporal) |
| ratio_pas_perf_tra       | Relation of number of Past Perfect tenses to all phrases in the srt                    | Syntactic (temporal) |
| pas_perf_prog_tra        | Number of Past Perfect Progressive tenses in the srt                                   | Syntactic (temporal) |
| ratio_pas_perf_prog_tra  | Relation of number of Past Perfect Progressive tenses to all phrases in the srt        | Syntactic (temporal) |
| will_tra                 | Number of Will-Future tenses in the srt  | Syntactic (temporal) |
| ratio_will_tra           | Relation of number of Will-Future tenses to all phrases in the srt                     | Syntactic (temporal) |
| fu_prog_tra              | Number of Future Progressive tenses in the srt   | Syntactic (temporal) |
| ratio_fu_prog_tra        | Relation of number of Future Progressive tenses to all phrases in the srt              | Syntactic (temporal) |
| fu_perf_tra              | Number of Future Perfect tenses in the srt   | Syntactic (temporal) |
| ratio_fu_perf_tra        | Relation of number of Future Perfect tenses to all phrases in the srt                  | Syntactic (temporal) |
| fu_perf_prog_tra         | Number of Future Perfect Progressive tenses in the srt                                 | Syntactic (temporal) |
| ratio_fu_perf_prog_tra   | Relation of number of Future Perfect Progressive tenses to all phrases in the srt      | Syntactic (temporal) |
| cond_sim_tra             | Number of Conditional Simple tenses in the srt   | Syntactic (temporal) |
| ratio_cond_sim_tra       | Relation of number of Conditional Simple tenses to all phrases in the srt              | Syntactic (temporal) |
| cond_prog_tra            | Number of Conditional Progressive tenses in the srt                                    | Syntactic (temporal) |
| ratio_cond_prog_tra      | Relation of number of Conditional Progressive tenses to all phrases in the srt         | Syntactic (temporal) |
| cond_perf_tra            | Number of Conditional Perfect tenses in the srt  | Syntactic (temporal) |
| ratio_cond_perf_tra      | Relation of number of Conditional Perfect tenses to all phrases in the srt             | Syntactic (temporal) |
| cond_perf_prog_tra       | Number of Conditional Perfect Progressive tenses in the srt                            | Syntactic (temporal) |
| ratio_cond_perf_prog_tra | Relation of number of Conditional Perfect Progressive tenses to all phrases in the srt | Syntactic (temporal) |
| gerund_tra               | Number of Gerund/Present Participle tenses in the srt                                  | Syntactic (temporal) |

Continuation on the next page

**Table 1 – Continuation**

| <b>Feature</b>      | <b>Description</b>   | <b>Category</b>      |
|---------------------|--|----------------------|
| ratio_gerund_tra    | Relation of number of Gerund/Present Participle tenses to all phrases in the srt | Syntactic (temporal) |
| perf_part_tra       | Number of Perfect Participle tenses in the srt                                   | Syntactic (temporal) |
| ratio_perf_part_tra | Relation of number of Perfect Participle tenses to all phrases in the srt        | Syntactic (temporal) |
| inf_tra             | Number of Present Infinitive tenses in the srt                                   | Syntactic (temporal) |
| ratio_inf_tra       | Relation of number of Present Infinitive tenses to all phrases in the srt        | Syntactic (temporal) |
| perf_inf_tra        | Number of Perfect Infinitive tenses in the srt                                   | Syntactic (temporal) |
| ratio_perf_inf_tra  | Relation of number of Perfect Infinitive tenses to all phrases in the srt        | Syntactic (temporal) |
| active_tra          | Number of active verb forms in the srt   | Syntactic (temporal) |
| ratio_active_tra    | Relation of number of active verb forms to all phrases in the srt                | Syntactic (temporal) |
| passive_tra         | Number of passive verb forms in the srt  | Syntactic (temporal) |
| ratio_passive_tra   | Relation of number of passive verb forms to all phrases in the srt               | Syntactic (temporal) |
| ADJP_sli            | Number of adjective phrases in the slides  | Syntactic (phrases)  |
| ratio_ADJP_sli      | Relation of number of adjective phrases to all phrases in the slides             | Syntactic (phrases)  |
| avg_ADJP_sli        | Average number of adjective phrases per line in the slides                       | Syntactic (phrases)  |
| ADVP_sli            | Number of adverb phrases in the slides   | Syntactic (phrases)  |
| ratio_ADVP_sli      | Relation of number of adverb phrases to all phrases in the slides                | Syntactic (phrases)  |
| avg_ADVP_sli        | Average number of adverb phrases per line in the slides                          | Syntactic (phrases)  |
| NP_sli              | Number of noun phrases in the slides   | Syntactic (phrases)  |
| ratio_NP_sli        | Relation of number of noun phrases to all phrases in the slides                  | Syntactic (phrases)  |
| avg_NP_sli          | Average number of noun phrases per line in the slides                            | Syntactic (phrases)  |
| PP_sli              | Number of prepositional phrases in the slides                                    | Syntactic (phrases)  |
| ratio_PP_sli        | Relation of number of prepositional phrases to all phrases in the slides         | Syntactic (phrases)  |
| avg_PP_sli          | Average number of prepositional phrases per line in the slides                   | Syntactic (phrases)  |
| S_sli               | Number of simple declarative clauses in the slides                               | Syntactic (phrases)  |
| ratio_S_sli         | Relation of number of simple declarative clauses to all phrases in the slides    | Syntactic (phrases)  |
| avg_S_sli           | Average number of simple declarative clauses per line in the slides              | Syntactic (phrases)  |
| FRAG_sli            | Number of fragments in the slides  | Syntactic (phrases)  |

Continuation on the next page



Table 1 – Continuation

| Feature          | Description  | Category            |
|------------------|--|---------------------|
| ratio_FRAG_sli   | Relation of number of fragments to all phrases in the slides   | Syntactic (phrases) |
| avg_FRAG_sli     | Average number of fragments per line in the slides   | Syntactic (phrases) |
| SBAR_sli         | Number of subordinate clauses in the slides  | Syntactic (phrases) |
| ratio_SBAR_sli   | Relation of number of subordinate clauses to all phrases in the slides   | Syntactic (phrases) |
| avg_SBAR_sli     | Average number of subordinate clauses per line in the slides   | Syntactic (phrases) |
| SBARQ_sli        | Number of direct questions with wh-element in the slides   | Syntactic (phrases) |
| ratio_SBARQ_sli  | Relation of number of direct questions with wh-element to all phrases in the slides                                  | Syntactic (phrases) |
| avg_SBARQ_sli    | Average number of direct questions with wh-element per line in the slides  | Syntactic (phrases) |
| SINV_sli         | Number of declarative sentence with subject-aux inversion in the slides  | Syntactic (phrases) |
| ratio_SINV_sli   | Relation of number of declarative sentence with subject-aux inversion to all phrases in the slides                   | Syntactic (phrases) |
| avg_SINV_sli     | Average number of declarative sentence with subject-aux inversion per line in the slides                             | Syntactic (phrases) |
| SQ_sli           | Number of yes/no questions and subconstituent of SBARQ excluding wh-element in the slides                            | Syntactic (phrases) |
| ratio_SQ_sli     | Relation of number of yes/no questions and subconstituent of SBARQ excluding wh-element to all phrases in the slides | Syntactic (phrases) |
| avg_SQ_sli       | Average number of yes/no questions and subconstituent of SBARQ excluding wh-element per line in the slides           | Syntactic (phrases) |
| VP_sli           | Number of verb phrases in the slides   | Syntactic (phrases) |
| ratio_VP_sli     | Relation of number of verb phrases to all phrases in the slides  | Syntactic (phrases) |
| avg_VP_sli       | Average number of verb phrases per line in the slides  | Syntactic (phrases) |
| WHADVP_sli       | Number of wh-adverb phrases in the slides  | Syntactic (phrases) |
| ratio_WHADVP_sli | Relation of number of wh-adverb phrases to all phrases in the slides   | Syntactic (phrases) |
| avg_WHADVP_sli   | Average number of wh-adverb phrases per line in the slides   | Syntactic (phrases) |
| WHNP_sli         | Number of wh-noun phrases in the slides  | Syntactic (phrases) |
| ratio_WHNP_sli   | Relation of number of wh-noun phrases to all phrases in the slides   | Syntactic (phrases) |
| avg_WHNP_sli     | Average number of wh-noun phrases per line in the slides   | Syntactic (phrases) |
| WHPP_sli         | Number of wh-prepositional phrases in the slides   | Syntactic (phrases) |
| ratio_WHPP_sli   | Relation of number of wh-prepositional phrases to all phrases in the slides  | Syntactic (phrases) |

Continuation on the next page

Table 1 – Continuation

| Feature                | Description  | Category            |
|------------------------|--|---------------------|
| avg_WHPP_sli           | Average number of wh-prepositional phrases per line in the slides                | Syntactic (phrases) |
| avg_phrases_sli        | Average number of phrases per line in the slides                                 | Syntactic (phrases) |
| ADJP_tra               | Number of adjective phrases in the srt   | Syntactic (phrases) |
| ratio_ADJP_tra         | Relation of number of adjective phrases to all phrases in the srt                | Syntactic (phrases) |
| avg_ADJP_tra           | Average number of adjective phrases per sentence in the srt                      | Syntactic (phrases) |
| ADVP_tra               | Number of adverb phrases in the srt  | Syntactic (phrases) |
| ratio_ADVP_tra         | Relation of number of adverb phrases to all phrases in the srt                   | Syntactic (phrases) |
| avg_ADVP_tra           | Average number of adverb phrases per sentence in the srt                         | Syntactic (phrases) |
| NP_tra                 | Number of noun phrases in the srt  | Syntactic (phrases) |
| ratio_NP_tra           | Relation of number of noun phrases to all phrases in the srt                     | Syntactic (phrases) |
| avg_NP_tra             | Average number of noun phrases per sentence in the srt                           | Syntactic (phrases) |
| PP_tra                 | Number of prepositional phrases in the srt                                       | Syntactic (phrases) |
| ratio_PP_tra           | Relation of number of prepositional phrases to all phrases in the srt            | Syntactic (phrases) |
| avg_PP_tra             | Average number of prepositional phrases per sentence in the srt                  | Syntactic (phrases) |
| S_tra                  | Number of simple declarative clauses in the srt                                  | Syntactic (phrases) |
| ratio_S_tra            | Relation of number of simple declarative clauses to all phrases in the srt       | Syntactic (phrases) |
| avg_S_tra              | Average number of simple declarative clauses per sentence in the srt             | Syntactic (phrases) |
| FRAG_tra               | Number of fragments in the srt   | Syntactic (phrases) |
| ratio_FRAG_tra         | Relation of number of fragments to all phrases in the srt                        | Syntactic (phrases) |
| avg_FRAG_tra, SBAR_tra | Average number of fragments per sentence in the srt                              | Syntactic (phrases) |
| SBAR_tra               | Number of subordinate clauses in the srt   | Syntactic (phrases) |
| ratio_SBAR_tra         | Relation of number of subordinate clauses to all phrases in the srt              | Syntactic (phrases) |
| avg_SBAR_tra           | Average number of subordinate clauses per sentence in the srt                    | Syntactic (phrases) |
| SBARQ_tra              | Number of direct questions with wh-element in the srt                            | Syntactic (phrases) |
| ratio_SBARQ_tra        | Relation of number of direct questions with wh-element to all phrases in the srt | Syntactic (phrases) |
| avg_SBARQ_tra          | Average number of direct questions with wh-element per sentence in the srt       | Syntactic (phrases) |
| SINV_tra               | Number of declarative sentence with subject-aux inversion in the srt             | Syntactic (phrases) |

Continuation on the next page

Table 1 – Continuation

| Feature              | Description   | Category            |
|----------------------|---|---------------------|
| ratio_SINV_tra       | Relation of number of declarative sentence with subject-aux inversion to all phrases in the srt                   | Syntactic (phrases) |
| avg_SINV_tra         | Average number of declarative sentence with subject-aux inversion per sentence in the srt                         | Syntactic (phrases) |
| SQ_tra               | Number of yes/no questions and subconstituent of SBARQ excluding wh-element in the srt                            | Syntactic (phrases) |
| ratio_SQ_tra         | Relation of number of yes/no questions and subconstituent of SBARQ excluding wh-element to all phrases in the srt | Syntactic (phrases) |
| avg_SQ_tra           | Average number of yes/no questions and subconstituent of SBARQ excluding wh-element per sentence in the srt       | Syntactic (phrases) |
| VP_tra               | Number of verb phrases in the srt   | Syntactic (phrases) |
| ratio_VP_tra         | Relation of number of verb phrases to all phrases in the srt  | Syntactic (phrases) |
| avg_VP_tra           | Average number of verb phrases per sentence in the srt  | Syntactic (phrases) |
| WHADVP_tra           | Number of wh-adverb phrases in the srt  | Syntactic (phrases) |
| ratio_WHADVP_tra     | Relation of number of wh-adverb phrases to all phrases in the srt   | Syntactic (phrases) |
| avg_WHADVP_tra       | Average number of wh-adverb phrases per sentence in the srt   | Syntactic (phrases) |
| WHNP_tra             | Number of wh-noun phrases in the srt  | Syntactic (phrases) |
| ratio_WHNP_tra       | Relation of number of wh-noun phrases to all phrases in the srt   | Syntactic (phrases) |
| avg_WHNP_tra         | Average number of wh-noun phrases per sentence in the srt   | Syntactic (phrases) |
| WHPP_tra             | Number of wh-prepositional phrases in the srt   | Syntactic (phrases) |
| ratio_WHPP_tra       | Relation of number of wh-prepositional phrases to all phrases in the srt  | Syntactic (phrases) |
| avg_WHPP_tra         | Average number of wh-prepositional phrases per sentence in the srt  | Syntactic (phrases) |
| avg_phrases_tra      | Average number of phrases per sentence in the srt   | Syntactic (phrases) |
| avg_trigram_sli      | Average number of trigrams per line in the slides   | Syntactic (other)   |
| avg_tetragram_sli    | Average number of tetragrams per line in the slides   | Syntactic (other)   |
| amount_statement_sli | Number of statements in the slides  | Syntactic (other)   |
| ratio_statement_sli  | Relation of number of statements to all lines in the slides   | Syntactic (other)   |
| amount_question_sli  | Number of questions in the slides   | Syntactic (other)   |
| ratio_question_sli   | Relation of number of statements to all lines in the slides   | Syntactic (other)   |
| amount_tok_sli       | Number of words in the slides   | Syntactic (other)   |

Continuation on the next page

**Table 1 – Continuation**

| <b>Feature</b>       | <b>Description</b>   | <b>Category</b>      |
|----------------------|--|----------------------|
| sum_tok_len_sli      | Number of characters in the slides                           | Syntactic<br>(other) |
| min_tok_len_sli      | Size of the shortest word in the slides                      | Syntactic<br>(other) |
| avg_tok_len_sli      | Average size of the words in the slides                      | Syntactic<br>(other) |
| max_tok_len_sli      | Size of the longest word in the slides                       | Syntactic<br>(other) |
| avg_trigram_tra      | Average number of trigrams per sentence in the srt           | Syntactic<br>(other) |
| avg_tetragram_tra    | Average number of tetragrams per sentence in the srt         | Syntactic<br>(other) |
| amount_statement_tra | Number of statements in the srt                              | Syntactic<br>(other) |
| ratio_statement_tra  | Relation of number of statements to all sentences in the srt | Syntactic<br>(other) |
| amount_question_tra  | Number of questions in the srt                               | Syntactic<br>(other) |
| ratio_question_tra   | Relation of number of statements to all sentences in the srt | Syntactic<br>(other) |
| amount_tok_tra       | Number of words in the srt                                   | Syntactic<br>(other) |
| sum_tok_len_tra      | Number of characters in the srt                              | Syntactic<br>(other) |
| min_tok_len_tra      | Size of the shortest word in the srt                         | Syntactic<br>(other) |
| avg_tok_len_tra      | Average size of the words in the srt                         | Syntactic<br>(other) |
| max_tok_len_tra      | Size of the longest word in the srt                          | Syntactic<br>(other) |
| flesch_ease_sli      | Flesch-Reading-Ease result of the slides                     | Readability          |
| flesch_kin_sli       | Flesch-Kincaid result of the slides                          | Readability          |
| gunning_fog_sli      | Gunning-Fog Index result of the slides                       | Readability          |
| smog_sli             | SMOG result of the slides                                    | Readability          |
| ari_sli              | Automated Readability Index result of the slides             | Readability          |
| coleman_sli          | Coleman-Liau Index result of the slides                      | Readability          |
| read_time_sli        | Reading time of the text of the slides                       | Readability          |
| flesch_ease_tra      | Flesch-Reading-Ease result of the srt                        | Readability          |
| flesch_kin_tra       | Flesch-Kincaid result of the srt                             | Readability          |
| gunning_fog_tra      | Gunning-Fog Index result of the srt                          | Readability          |
| smog_tra             | SMOG result of the srt                                       | Readability          |
| ari_tra              | Automated Readability Index result of the srt                | Readability          |
| coleman_tra          | Coleman-Liau Index result of the srt                         | Readability          |
| read_time_tra        | Reading time of the text of the srt                          | Readability          |
| avg_freq_tok_sli     | Average frequency of a word in the slides                    | Lexical              |
| min_age_sli          | Minimum age of acquisition of a word in the slides           | Lexical              |
| avg_age_sli          | Average age of acquisition of a word in the slides           | Lexical              |
| max_age_sli          | Maximum age of acquisition of a word in the slides           | Lexical              |
| amount_syl_sli       | Number of syllables in the slides                            | Lexical              |
| amount_one_syl_sli   | Number of syllables in the slides                            | Lexical              |

Continuation on the next page

**Table 1 – Continuation**

| <b>Feature</b>       | <b>Description</b>  | <b>Category</b>     |
|----------------------|---|---------------------|
| amount_two_syl_sli   | Number of syllables in the slides   | Lexical             |
| amount_psyl_sli      | Number of syllables in the slides   | Lexical             |
| amount_hard_sli      | Number of syllables in the slides   | Lexical             |
| avg_syl_sli          | Average number of syllables per word in the slides                              | Lexical             |
| ratio_one_syl_sli    | Relation of number of monosyllabic words to all unfiltered tokens in the slides | Lexical             |
| ratio_two_syl_sli    | Relation of number of two-syllable words to all unfiltered tokens in the slides | Lexical             |
| ratio_psyl_sli       | Relation of number of polysyllabic words to all unfiltered tokens in the slides | Lexical             |
| ratio_hard_sli       | Relation of number of hard words to all unfiltered tokens in the slides         | Lexical             |
| amount_uni_tok_sli   | Number of unique words in the slides  | Lexical             |
| ratio_uni_tok_sli    | Relation of number of unique words to all words in the slides                   | Lexical             |
| amount_uni_lemma_sli | Number of unique lemmas of the words in the slides                              | Lexical             |
| ratio_uni_lemma_sli  | Relation of number of unique lemmas of the words to all words in the slides     | Lexical             |
| avg_freq_tok_tra     | Average frequency of a word in the srt  | Lexical             |
| min_age_tra          | Minimum age of acquisition of a word in the srt                                 | Lexical             |
| avg_age_tra          | Average age of acquisition of a word in the srt                                 | Lexical             |
| max_age_tra          | Maximum age of acquisition of a word in the srt                                 | Lexical             |
| amount_syl_tra       | Number of syllables in the srt  | Lexical             |
| amount_one_syl_tra   | Number of syllables in the srt  | Lexical             |
| amount_two_syl_tra   | Number of syllables in the srt  | Lexical             |
| amount_psyl_tra      | Number of syllables in the srt  | Lexical             |
| amount_hard_tra      | Number of syllables in the srt  | Lexical             |
| avg_syl_tra          | Average number of syllables per word in the srt                                 | Lexical             |
| ratio_one_syl_tra    | Relation of number of monosyllabic words to all unfiltered tokens in the srt    | Lexical             |
| ratio_two_syl_tra    | Relation of number of two-syllable words to all unfiltered tokens in the srt    | Lexical             |
| ratio_psyl_tra       | Relation of number of polysyllabic words to all unfiltered tokens in the srt    | Lexical             |
| ratio_hard_tra       | Relation of number of hard words to all unfiltered tokens in the srt            | Lexical             |
| amount_uni_tok_tra   | Number of unique words in the srt   | Lexical             |
| ratio_uni_tok_tra    | Relation of number of unique words to all words in the srt                      | Lexical             |
| amount_uni_lemma_tra | Number of unique lemmas of the words in the srt                                 | Lexical             |
| ratio_uni_lemma_tra  | Relation of number of unique lemmas of the words to all words in the srt        | Lexical             |
| sum_lines            | Number of lines in the slides   | Structural (slides) |
| min_line_len         | Minimum number of words per line in the slides                                  | Structural (slides) |
| avg_line_len         | Average number of words per line in the slides                                  | Structural (slides) |

Continuation on the next page

Table 1 – Continuation

| Feature            | Description  | Category            |
|--------------------|--|---------------------|
| max_line_len       | Maximum number of words per line in the slides             | Structural (slides) |
| min_lines          | Minimum number of lines per slide                          | Structural (slides) |
| avg_lines          | Average number of lines per slide                          | Structural (slides) |
| max_lines          | Maximum number of lines per slide                          | Structural (slides) |
| min_words_slide    | Minimum number of words per slide                          | Structural (slides) |
| avg_words_slide    | Average number of words per slide                          | Structural (slides) |
| max_words_slide    | Maximum number of words per slide                          | Structural (slides) |
| min_line_chars     | Minimum number of characters per line in the slides        | Structural (slides) |
| avg_line_chars     | Average number of characters per line in the slides        | Structural (slides) |
| max_line_chars     | Maximum number of characters per line in the slides        | Structural (slides) |
| amount_slides      | Amount of slides of a presentation                         | Structural (slides) |
| amount_subtitles   | Number of subtitles in the srt                             | Structural (srt)    |
| amount_sentences   | Number of sentences in the srt                             | Structural (srt)    |
| speak_time         | Subtitle display time in the srt                           | Structural (srt)    |
| speak_difference   | Difference between display time and reading time (180 WPM) | Structural (srt)    |
| min_sen_len        | Minimum number of words per sentence in the srt            | Structural (srt)    |
| avg_sen_len        | Average number of words per sentence in the srt            | Structural (srt)    |
| max_sen_len        | Maximum number of words per sentence in the srt            | Structural (srt)    |
| min_sen_chars      | Minimum number of characters per sentence in the srt       | Structural (srt)    |
| avg_sen_chars      | Average number of characters per sentence in the srt       | Structural (srt)    |
| max_sen_chars      | Maximum number of characters per sentence in the srt       | Structural (srt)    |
| embed_slide        | Average sentence embedding of the slides                   | Semantic            |
| similarity_sli     | Average similarity between the lines in the slides         | Semantic            |
| embed_srt          | Average sentence embedding of the srt                      | Semantic            |
| similarity_tra     | Average similarity between the sentences in the srt        | Semantic            |
| diff_similarity    | Average similarity between the sentences in the srt        | Semantic            |
| similarity_vectors | similarity of the two embeddings                           | Semantic            |
| Person_ID          | The ID of a user   | User specific       |