

**Table S1.** Overview of current studies with NLP Input features and Optimal Predictor features specification.

Author (year)	Linguistic analysis method	NLP Input features	Optimal predictors
Yokoi et al. [45]	MeCab Cabocha	<p><u>Phonetic and Phonological</u>: filler ratio, filler utterance ratio.</p> <p><u>Lexico-semantic</u>: noun ratio, verb ratio, adjective ratio, adjectival noun ratio, auxiliary verb ratio, particle ratio, adnominal adjective ratio, adverb ratio, conjunction ratio, interjection ratio, symbol ratio, other's ratio, pronoun ratio, general noun ratio, proprietary noun ratio, pronoun ratio (dispersion), general noun ratio (dispersion), proprietary noun ratio (dispersion), prefix ratio, pronoun utterance ratio, general noun utterance ratio, proprietary noun utterance ratio, pronoun utterance ratio, general noun utterance ratio, proprietary noun utterance ratio, case particle utterance ratio, verb utterance ratio, adverb utterance ratio, adnominal adjective utterance ratio, interjection utterance ratio, TTR, different noun ratio, vocabulary level.</p> <p><u>Morphosyntactic</u>: case particle ratio, case particle ratio (dispersion), maximum dependency distance, dependency depth, the average number of dependencies.</p>	<p><u>Phonetic and Phonological</u>: filler utterance ratio</p> <p><u>Lexico-semantic</u>: verb ratio, general noun utterance ratio, proper noun utterance ratio, verb utterance ratio.</p> <p><u>Morphosyntactic</u>: case particle ratio (dispersion)</p>
Escobar-Grisales et al. (a) [43]	BETO	<u>NLP extracted features</u> : verb embeddings, noun embeddings, embeddings statistical functionals (mean, standard deviation, skewness, kurtosis).	<u>NLP extracted features</u> : verb embeddings
Escobar-Grisales et al. (b) [44]	W2V BERT BETO	<u>NLP extracted features</u> : embeddings statistical functionals (mean, standard deviation, skewness, kurtosis), word embeddings.	<u>NLP extracted features</u> : word embeddings.
Favaro et al. [46]	Whisper+ SpaCy	<p><u>Lexico-semantic</u>: nouns, verbs, adjectives, numerals, auxiliaries.</p> <p><u>Morphosyntactic</u>: mean length words, utterances number, MLU in words, number of noun phrases, number of verb phrases, number of prepositional phrases.</p> <p><u>Discourse and Pragmatic</u>: total words, words number with no function words.</p>	n/a
Garcia et al. [47]	LSA+POS tag GloVeV GloVeOS	<u>NLP extracted features</u> : p-RSF score, word embeddings, verb embeddings.	<u>NLP extracted features</u> : verb embeddings
Eyigoz et al. [42]	Freeling Morphodita	<u>Lexico-semantic</u> : verb ratio, noun ratio, adjective ratio, adverb ratio, pronoun ratio, determiner ratio, conjunction ratio, adposition ratio, interjection ratio, pronoun type, verb	<p><i>Spanish</i></p> <p><u>Lexico-semantic</u>: subordinating conjunctions, proper nouns</p>

		<p>type, determiner type, noun type, conjunction type, adverb type, adposition type, punctuation.</p> <p><u>Morphosyntactic</u>: pronoun gender, pronoun number, pronoun case, verb mood, verb tense, verb person, verb number, determiner gender, determiner number, noun gender, noun number, noun degree.</p> <p><u>Morphological consistency</u>: POS Probability scores, Statistical functionals of POS probability score (mean, standard deviation, minimum, maximum, skewness, kurtosis).</p>	<p><u>Morphological consistency</u>: proper nouns skewness, Present tense verbs mean</p>
			<p><i>German</i></p> <p><u>Morphological consistency</u>: neuter gender pronouns kurtosis, Verb person not specified skewness, Determiner in accusative case skewness, Feminine nouns standard deviation</p>
			<p><i>Czech</i></p> <p><u>Lexico-semantic</u>: personal pronoun.</p> <p><u>Morphological consistency</u>: person not specified skewness, 2nd most frequent variant kurtosis, Masculine gender skewness.</p>
Perez-Toro et al. [41]	BoW TF-IDF W2V	<p><u>Lexico-semantic</u>: words frequency</p> <p><u>NLP extracted features</u>: word or combination of words relative frequency, word embeddings</p>	<u>Lexico-semantic</u> : words frequency.
Eyigoz et al. [40]	Morfessor	<u>Lexico-semantic</u> : words frequency, morpheme suffixes, morpheme stems, morpheme prefixes, morpheme lengths, words length.	<u>Lexico-semantic</u> prefix-prefix pair, prefix length probability, prefixes probability, prefix-stem probability.
Jessiman et al. [39]	SpaCy So-Cal GloVe Cosine-similarity	<p><u>Lexico-semantic</u>: familiarity, TTR for dependency relations, words frequency, POS TTR, TTR, concreteness, imageability, typical age of acquisition, frequency in everyday life.</p> <p><u>Morphosyntactic</u>: MLU in words, MLU, maximum verb dependency distance, maximum word dependency distance, sparse bag-of-relations features.</p> <p><u>Sentiment</u>: sentiment score.</p> <p><u>NLP extracted features</u>: word embeddings, embeddings similarity.</p>	n/a
Garcia et al. [25]	LSA	<u>NLP extracted features</u> : NLP-Semantic	<u>NLP Extracted features</u> : NLP Semantics (n=4).
	Stuttgart TreeTagger	<u>Lexico-semantic</u> : subordinating conjunctions, negative markers, proper nouns, inter-sentential separation markers	<u>Lexico-semantic</u> : subordinating conjunctions, negative markers.

*Note. NLP= Natural Language Processing; MeCab= Japanese Morphological analysis implementation; W2V= Word2Vec; LSA= Latent Semantic Analysis; GloVe= Global Vectors for Word Representation; GloVeV= Global Vectors for Word Representation verb-to-verb semantic distance; GloVeOS= Global Vectors for Word Representation overall semantic structure; BERT= Bidirectional Encoder Representations from Transformers; BETO= Spanish-BERT; MLU= Mean Length Unit; BoW= Bag-of-Words; TF= Term Frequency; IDF= Inverse Document Frequency; SO-Cal= Semantic Orientation Calculator; POS= Part of Speech; P-RSF=Proximity-to-Reference-Semantic-Field; TTR= type-token ratio; n/a= not available,*