# Significant paper presentation

## Computationally Detecting and Quantifying the Degree of Bias in Sentence-Level Text of News Stories

**Published in:** HUSO 2015: The First International Conference on Human and Social Analytics

**Authors: C.J. Hutto, Dennis Folds, Scott Appling**

**Presented by Aayushi, Anjali, Sneha, Vaibhav(Group 13) for Cmpe 255, fall 2019**

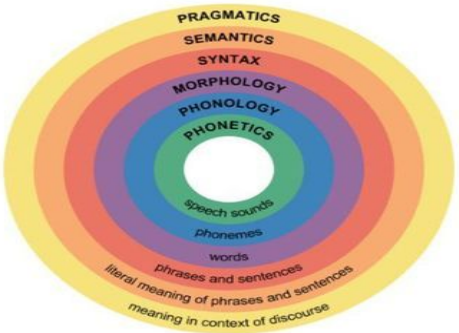
# Introduction to Linguistics

## Previously on introduction to **Linguistics**...

The scientific study of human language

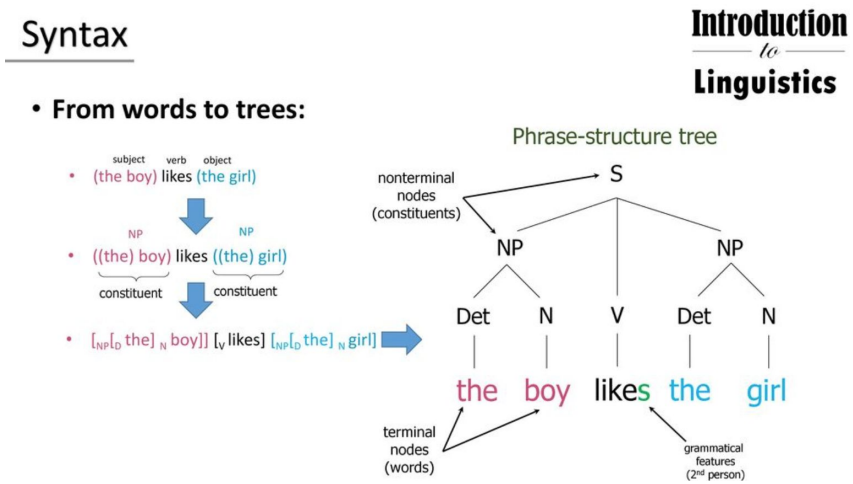
- **Phonetics** (physical nature of speech)
- **Phonology** (use of sounds in language)
- **Morphology** (the structure of words)
- **Syntax** (sentence structure)
- **Semantics** (meaning of words & how they combine into sentences)
- **Pragmatics** (the study of language use and communication in context)

**Introduction** *to* **Linguistics**

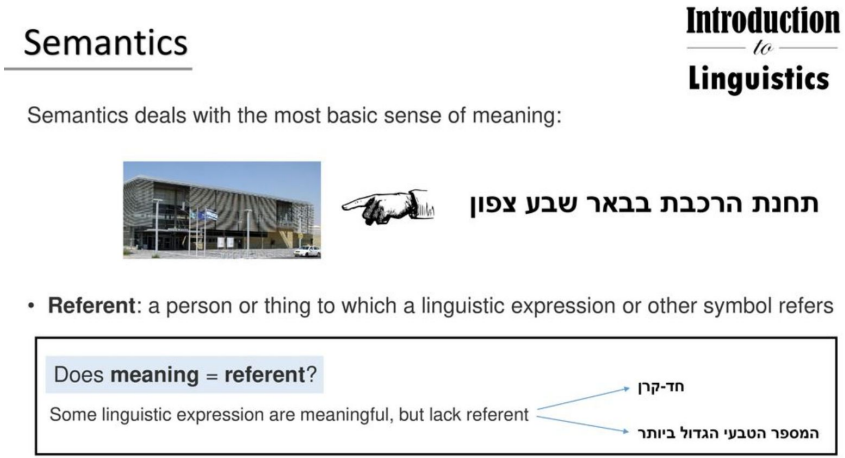


**The features that the paper discusses of biased language can be broadly divided as follows:**

1) **Structural analysis at the sentence level -** In this we observe characteristics of the text statement as a whole, considering syntactical, grammatical, and structural properties. This consists of Sentiment score, Subjectivity, Modality, Mood and Readability of the text.



**2) Linguistic analysis at the sentence level -** This aims at detecting either epistemological bias(knowledge contained in the text) or framing bias(avoiding risks while writing the review). Includes features such as Factive verbs, Implicative verbs, Assertive verbs, Hedges, Strong subjective intensifiers, weak subjective intensifiers, bias terms, Opinion words, Degree Modifiers, Coherence Markers, Causation words, Certainty words, Tentative words, 3rd person pronoun, Achievement words, Work words, Discrepancy words, Conjunctions, prepositions, Adverbs and Auxiliary verbs.



# Preliminary results using Linear regression model

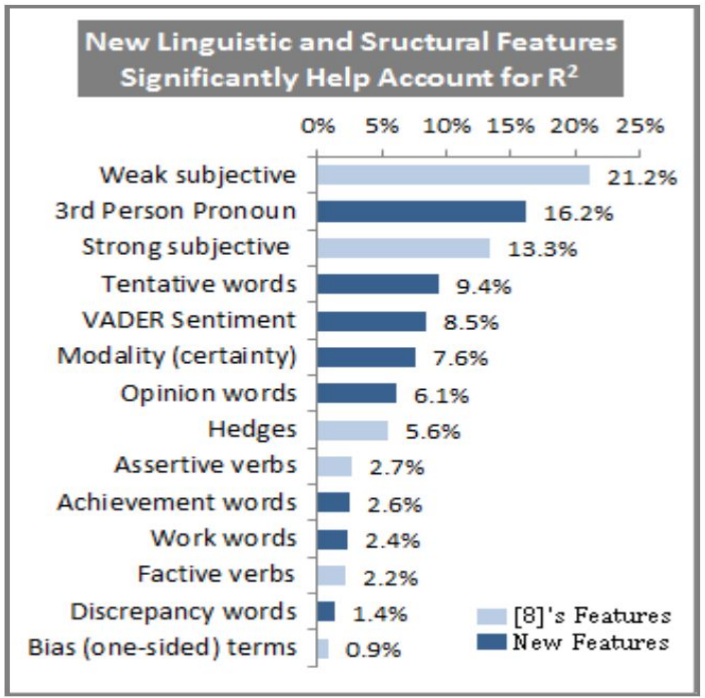



Figure 1. Proportion of variance accounted for by each feature in the improved model using the mean $R^2$ of three regression techniques (feature added to model first, feature added to model last, and feature beta squared).

TABLE II: COEFFICIENTS, ERROR, T-VALUES, AND *P*-VALUES FOR THE IMPROVED MODEL. $F(14,26) = 11.3$, $P = 1.04E-07$.

| | *b* | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -0.56 | 0.19 | -3.02 | 0.006 |
| **Strong subjective** | 5.10 | 1.07 | 4.74 | 0.000*** |
| **3rd Person Pronoun** | 8.36 | 1.95 | 4.30 | 0.000*** |
| **Weak subjective** | 4.87 | 1.19 | 4.08 | 0.000*** |
| **Modality (certainty)** | 0.52 | 0.15 | 3.42 | 0.002** |
| **VADER Sentiment** | 0.35 | 0.11 | 3.13 | 0.004** |
| **Tentative words** | 4.60 | 1.65 | 2.79 | 0.010** |
| **Opinion words** | -2.05 | 0.95 | -2.16 | 0.040* |
| **Achievement words** | 5.74 | 2.66 | 2.16 | 0.040* |
| **Factive verbs** | -16.64 | 8.39 | -1.98 | 0.058` |
| **Work words** | 9.81 | 5.20 | 1.89 | 0.070` |
| **Hedges** | 3.06 | 1.75 | 1.75 | 0.092` |
| **Assertive verbs** | -3.58 | 2.16 | -1.66 | 0.110 |
| **Discrepancy words** | 5.66 | 3.62 | 1.56 | 0.130 |
| **Bias (one-sided) terms** | -0.95 | 0.74 | -1.30 | 0.206 |

Signif. level codes: $p < 0.001$*** $p < 0.01$** $p < 0.05$* $p < 0.1$`

# Conclusion & References

- This paper helped us identify features for our model for detecting biased/fake reviews.
- We have used features like ‘number of Nouns’, ‘number of adjectives’, ‘polarity of the text reviews’, etc as our features based on this paper.
- References: https://slideplayer.com/slide/12566202/

  https://pdfs.semanticscholar.org/3722/40da2a416abfb406426af71f084c988ff7d9.pdf