

Summary

VARIATION IN THE CONTEXTUALITY OF LANGUAGE: AN EMPIRICAL MEASURE

Submitted by:
Pallavit Aggarwal
Intelligent Systems – 22333721

In this paper the authors discuss a new way of analyzing the contextuality of natural language in order to make it less ambiguous and more empirical. It starts off with defining the low-context and high-context situations and goes on to describe how different situations involve the use of different determiners and words and with that different degree of context, for example an informal conversation might assume a high degree of implicit context whereas a formal conversation like that in legal or office setup might require formal statements and low context-dependent statements with a high degree of information available in that.

The paper first examines these different linguistic communications and then provides an empirical formula to estimate the degree of formalness or the degree of context provided in a sentence.

In Formality vs Contextuality, ambiguity in natural language and the need to determine context and consequently the scope of the definition of context is established. The difference between deep formal statements and surface formal statements are also highlighted though not discussed in depth in the later parts of the papers where analysis of different corpus is taken.

In the next section, Measuring language contextuality, different considerations are discussed. Situations where context and the knowledge of historical conversation and other factors are to be given due importance in the proper evaluation of a text. Moreover, methodologies used by social scientists and “bootstrapping” procedure is discussed and the final approach used by the authors has developed by bootstrapping initial thermotical concepts together alongside different observations and studies that they refer to.

The approach used by the authors is to first divide the words into more context-dependent or less context-dependent class. Next deictic and non-deictic words are analyzed. Non-deictic are nouns, adjectives, prepositions and articles whose frequency is expected to increase with formality of a text and non-deictic words like pronouns, verbs, adverbs and interjections usually increase with the contextuality of a conversation i.e. when it becomes less formal. A basic arithmetic of adding up the frequencies of all formal categories and subtracting the frequencies of contextual categories and finally normalizing the score to be out of 100, gives the F-Measure or the measure of formalness in sentences. The more formal the language excerpt, the higher the value of F.

The F-measure is now tested on different corpora of Dutch and Italian texts. The findings corroborate with the initial hypotheses and also show that the written text is more formal than spoken conversations. The texts that are of scientific nature get the highest F-score owing to the formalness and the completeness of the texts in which they have to be presented. In the section about the universality of the Formality as a factor, the authors mention another study in support of their concept where using factor analysis, the same methodology emerges and the stylistic variations across different genres and themes ultimately boil down the analysis to

formal and contextual words and their usage in the excerpt. This F-measures authors discuss can be improved if further analysis on the correlation of the words is provided and then a formality score appointed to them instead of simply checking nouns and appointing them a formal score, as an example “Daddy” and “Thing” both are nouns but would actually need high-context to figure out the intention and usage behind these words and therefore should have a lower F-score assigned to them. Such optimizations can improve the overall calculations of the formality of an excerpt.

There are some non-linguistic determinants like situation, gender, introversion, level of education that deeply influence the quality of the language used and the usage of formal or informal words. A person with a higher level of education has a higher chance of stating things and forming sentences with a more detail thus providing it a more formal attribute and improving the F-score. Another example is that female speakers tend to use more contextual language than male speakers, and older speakers tend to use less contextual language than younger speakers. These are brought out as limitations to the approach and the authors acknowledge these external factors that should be modelled into the approach to make it more robust overall.

From a text analysis point of view, these findings have implications for our understanding of how language is used in different contexts, and for the development of natural language processing tools that can better capture the contextual variability of language.

References

1. Heylighen, F. & Dewaele, J.-M. (2002). Variation in the Contextuality of Language: An Empirical Measure. *Foundations of Science*, 7, 293–340.