

EPISTEMIC MODALITY IN COMPUTER-MEDIATED DISCOURSE:
THE CASE OF WIKIPEDIA'S ACADEMIC CORE

MUHAMMAD M. ABDUL-MAGEED

Department of Linguistics
Indiana University, Bloomington,
USA
mabdulma@indiana.edu

Epistemic modality serves to indicate how confident language users are about the truth of the ideational material they convey; it is often subdivided into two types of markers, hedges and boosters. Hedges are linguistic devices like *perhaps*, *I guess*, and *to a certain extent* that speakers employ to reduce the degree of liability or responsibility they might face in expressing the ideational material. Boosters are elements like *definitely*, *I assure that*, and *of course* that speakers use to emphasize what they do really believe—or would like their audiences to think they believe. Epistemic modality has been shown to play an especially significant role in written academic discourses, as scholars use these markers to create research spaces for themselves, construct evidence, politely refute others' claims, etc. As far as we know, however, no research has investigated the employment of epistemic modality in computer-mediated discourse (CMD). Since arguments about CMD have sometimes lacked empirical support (e.g., Crystal's [2001] characterization of 'Netspeak' as a third medium of language that is neither speech nor writing), empirical research is needed to target the various linguistic features of this type of communication and compare it with other modalities.

The current paper seeks to identify how it is that epistemic modality is employed in what I label 'the academic core' of Wikipedia, the free collaboratively-authored online encyclopedia. Since epistemic modality markers are frequent in offline academic discourses, they were expected to be of importance in Wikipedia's academic core. I hence employ a corpus-based approach to investigate the employment of epistemic modality in a 'snapshot' corpus of 1.5 million words of the encyclopedia's articles as well as their parallel Talk Pages (TPs) from 4 soft and 4 hard sciences. The corpus was broken down into 22 sub-corpora based on disciplinary as well as generic variations and the frequencies of the devices were retrieved automatically and a specific number of randomly selected cases was checked manually. Results show that the employment of these devices differed according to the genre as well as the discipline of each sub-corpus. Comparison to earlier literature on traditional academic discourse shows that the frequency of epistemic modality in Wikipedia's articles is closer to Ph.D. than MA theses and that disciplinary differences (i.e., between soft and hard sciences) in traditional academic discourses is traversed to the two naming spaces of Wikipedia (i.e., articles and TPs). I discuss the implications of these findings with reference to the status of Wikipedia, massively distributed collaboration, and existing rhetorical theory. Based on the results, I also refute some non-empirically supported propositions about CMD (e.g., Crystal's 2001 claim that 'Netspeak' is a unified phenomenon) and argue that my results provide empirical support for Herring's (2007) call for a more nuanced approach to the analysis of computer-mediated discourse.

References

- Crystal, D. (2001). *Language and the internet*. Cambridge: Cambridge University Press.
- Herring, S. C. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet*, 4(1). Retrieved March 17, 2008
<http://www.languageatinternet.de/articles/2007/761>.