

# Keynote: Profiling the Personality of Social Media Users

Walter Daelemans

CLiPS, University of Antwerp, The Netherlands

walter.daelemans@uantwerpen.be

In the social media, everybody is a writer, and many people freely give away their personal information (age, gender, location, education, and, often indirectly, also information about their psychology such as personality, emotions, depression etc.). By linking the text they write with this metadata of many social media users, we have access to large amounts of rich data about real language use. This makes possible the development of new applications based on machine learning, as well as a new empirical type of sociolinguistics based on big data.

In this paper I will provide a perspective on the state of the art in profiling social media users focusing on methods for personality assignment from text. Despite some successes, it is still uncertain whether this is even possible, but if it is, it will allow far-reaching applications. Personality is an important factor in life satisfaction and determines how we act, think and feel. Potential applications include targeted advertising, adaptive interfaces and robots, psychological diagnosis and forensics, human resource management, and research in literary science and social psychology.

I will describe the personality typology systems currently in use (MBTI, Big Five, Enneagram), the features and methods proposed for assigning personality, and the current state of the art, as witnessed from, for example, the PAN 2015 competition on profiling and other shared tasks on benchmark corpora. I will also go into the many problems in this subfield of profiling; for example the unreliability of the gold standard data, the shaky scientific basis of the personality typologies proposed, and the low accuracies achieved for many traits in many corpora. In addition, as is the case for the larger field of profiling, we are lacking sufficiently large balanced corpora for studying the interaction with topic and register, and the interactions between profile dimensions such as age and gender with personality.

As a first step toward a multilingual shared task on personality profiling, I will describe joint work with Ben Verhoeven and Barbara Plank on collecting and annotating the TwiSty corpus (Verhoeven et al., 2016). TwiSty (<http://www.clips.ua.ac.be/datasets/twisty-corpus>) contains personality (MBTI) and gender annotations for a total of 18,168 authors spanning six languages: Spanish, Portuguese, French, Dutch, Italian, German. A similar corpus also exists for English. It may be a first step in the direction of a balanced, multilingual, rich social media corpus for profiling.

## References

Ben Verhoeven, Walter Daelemans, and Barbara Plank. 2016. Twisty: A multilingual twitter stylometry corpus for gender and personality profiling. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. European Language Resources Association (ELRA).

## Bio

Walter Daelemans is professor of Computational Linguistics at the University of Antwerp where he directs the CLiPS computational linguistics research group. His research interests are in machine learning of natural language, computational psycholinguistics, computational stylometry, and language technology applications, especially biomedical information extraction and cybersecurity systems for social networks. He has supervised 25 finished PhDs and (co-)authored more than 300 publications. He was elected EURAI Fellow, ACL Fellow, and member of the Royal Academy for Dutch Language and Literature.