

PLIN 2016: “Language and the new (instant) media”

12 May 2016, Louvain-la-Neuve, Belgium.

Domain Keywords: Mediated discourse analysis, Normalisation, Natural Language Processing.

Medium Keywords: SMS.

Cédric Lopez, Mathieu Roche, Rachel Panckhurst

“Non-standard texts: from theoretical positions to Natural Language Processing normalisation”

[50 words]

Our digital resource of 88,000 anonymised French text messages, the *88milSMS* corpus, and sociolinguistic questionnaire data, are available (<http://88milsms.huma-num.fr>). Our theoretical position and Natural Language Processing (NLP) investigation techniques, including mediated discourse analysis on SMS-writing, ‘unknown’ item classification, alignment and normalisation methods, are envisaged for future implementation in real-life applications.