

Assuming identities online: description, development and ethical implications (AIO)

Professor Tim Grant (PI) and Dr Nicci MacLeod (RA)



Context

One area of undercover online police work in which linguists have been able to offer their expertise is in assisting police officers in the assumption of alternative online identities. Such a strategy might be considered in cases where a child has been identified as having been the victim of online grooming. In such a situation the child will be removed and an undercover officer (UCO) may be authorised to take on the identity of the child to continue the online interaction with a view to making an arrest. Our two-year ESRC-funded project, in collaboration with the University of Warwick, West Midlands Police, and 2LXG, aims to inform this process. This poster focuses on one narrow aspect of our project, namely cues to impersonation evident in our experimental tasks with students.

Research questions

- What are the linguistic sources of individual and group variation in style and what is the relationship between linguistic style and online identity performance?
- What linguistic analysis is necessary and sufficient to describe an online linguistic persona to the extent it could be successfully assumed by another individual, and with what level of accuracy and confidence can individuals detect the substitution of one interlocutor with another?

Experimental design

To answer these questions, a series of experiments was designed within which participants were tasked either with *impersonating* a specific individual, or *detecting* the substitution of one individual with another, over Instant Messaging (IM). Those who were tasked with *impersonation* ('Impersonators') had varying levels of preparation for the task, ranging from none at all up to several days to revise with their target's historical chat logs (Fig 1). Those who were asked to *detect* substitution ('Judges') were required to record their opinions on when the switch had taken place, their confidence in this opinion, and the criteria that led them to this decision (Fig. 2 & 3).

This experimental data was used to supplement the genuine and training data we accessed through our partnership with West Midlands Police.

Analysis

Assessing the extent to which particular features of an individual's language use are noticed by their interlocutor involved assigning each criterion mentioned by judges to a particular level of linguistic analysis, as set out by Herring (2004; 2012). The four domains are: structure (e.g. spelling, punctuation, syntax); meaning (e.g. speech acts); interaction (e.g. topic control, turn length) and social behaviour (e.g. identities, relationships).

The results for each level appear in Fig. 4.

23

The person you will be impersonating appears as **judge_participant_24**

interlocutor_participant_22_23: Hey there!

judge_participant_24: Hello

judge_participant_24: How are you?

interlocutor_participant_22_23: I'm pretty good thanks, really want a nap!

interlocutor_participant_22_23: How about yourself?

judge_participant_24: I'm good a bit tired but I think I manage haha

interlocutor_participant_22_23: We're all so tired lately

interlocutor_participant_22_23: Think it's the time of year

Handwritten notes: USUALLY CAPITALISE, CAPITALISE START OF SENTENCES & NAMES, HAHA POPULAR (7) (over 1/4 time), SHORTER REPLIES, FEWER/SLOWER REPLIES, STANDARD APOSTROPHES

Judges' Response Form

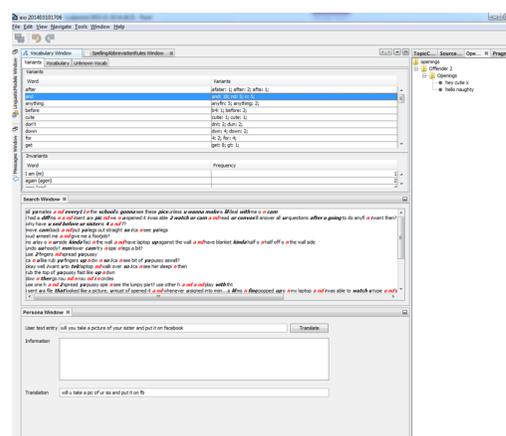
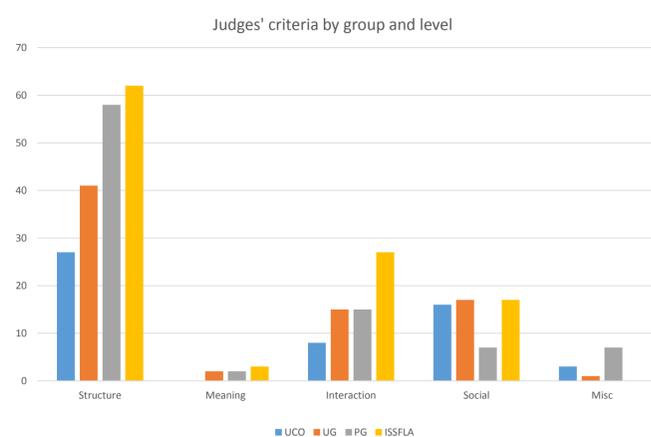
Condition : Homework

Participant Number 24

You have just taken part in an online chat, during which your chat partner was swapped for another individual. Please answer the following questions.

1. Please give the time of where you believe the switch occurred, with a brief summary of the content at that point.
14:44 - yes you are - speaking about different languages - there are 2 transmissions
2. How confident are you of this judgement (1 being least confident, 10 being most confident).
8
3. Please provide information about what leads you to believe a switch occurred at this point. This could relate to the language used, the topics discussed, the general tone, etc.
There are two transmissions and the writer has continued the sentence onto the next line. After this there are more chunks of speech whereas there were ~~wasnt~~ wasn't any before.

191	English proficiency; length of response; tone more enthusiastic; use of contractions.									
192	Much longer speech turns (phrases)									
193	Possible the first person was more upbeat?									
194	Change of use of apostrophe (acute vs. grave); typing speed quickened									
195	Suddenly there are fullstops after each sentence									
196	its - it's - it is; with a full stop; speed of typing/ replying									
197	speed of typing increases; use of American 'movies' for 'films'									
198	Interpretor - spelling mistake gave it away.									
199	More typos, change in style; 'nope' seemed more like interlocutor's style.									
200	Author A: emoticons in separate messages; always uses comma. Author B: emoticons at the end									



Findings

Data collected via the experiments present myriad opportunities for exploration. For example, comments provided by judges about what has 'tipped them off' to the possibility of substitution of their interlocutor provide real insight into the levels of linguistic analysis on which undercover operatives might want to focus. Features at the structural level – for example spelling and vocabulary choices, punctuation, and so on – are clearly the first cue to substitution, followed by features at the interactional level, such as turn length and speed of response (see Fig.4).

Judges' responses were coded as being at the 'social' level when they mentioned elements such as 'tone' and 'personality'. A selection of these appears below:

change of tone, and longer sentences, more sophisticated, less playful.

Possible the first person was more upbeat?

seemed easier to chat, more natural, more on same page

change in personality more towards the end.

Since comments at the social level make up a substantial proportion of the Judges' responses, we continue to examine how this can best be incorporated into the training.

Impact

These analyses have fed into the development of Identik, a software tool designed to assist UCOs with linguistic analysis and impersonation (see Fig. 5).

References

Herring, S.C., 2004. Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior. In S. A. Barab, R. Kling, & J. H. Gray, eds. *Designing for Virtual Communities in the Service of Learning*. Cambridge University Press, pp. 338–376.

Herring (2012) Discourse in Wwb 2.0: Familiar, Reconfigured, and Emergent. In D. Tannen & A. M. Tester (Eds.), *Georgetown University Round Table on Languages and Linguistics 2011: Discourse 2.0: Language and new media*. Washington, DC: Georgetown Univ. Press.