A corpus-based approach to informality: the case of Internet chat

Stefan Grondelaers – Dirk Speelman – Dirk Geeraerts
RU *Quantitative Lexicology and Variational Linguistics*
Dept. of Linguistics – University of Leuven
Blijde-Inkomststraat 21
B-3000 Leuven, Belgium
fax: +32 16 324767
e-mail: stefan.grondelaers@arts.kuleuven.ac.be

The language of IRC – Internet Relay Chat – is in many respects an example of "spoken language in written form": although produced in a written medium, it shares with spoken language a dialogical immediacy that ordinary written text usually lacks, as a result of which, it tends to appear highly informal, even to the untrained observer.

Linguists should not, however, be misguided by this global impression of informality: the language of IRC is informal in at least four different respects. In addition to its dialogical character (a *situational* form of informality which manifests itself in a.o. the relatively higher frequency of 2nd person pronouns and vocatives), IRC is characterized by an abundance of abbreviation and ellipsis, which are parameters of the production demands of a written *medium* whose users try to imitate spoken language (e.g. Hentschel 1998). A third source of informality in IRC is *speaker-related*: being for the most part pre-adult males, chatters tend to indulge significantly more often in tabooed topics and territorial behavior (Hentschel 1998, De Gryse 2000). A fourth group of informality-inducing factors is *register-related*: no matter how colloquial the situation and/or medium, chatters may choose to sound "vernacular", or to maintain a more formal standard, to "look after" their language.

Our paper has a methodological and an empirical aim. We will demonstrate, first, that whereas stylometric approaches based on the calculation of isolated variables (e.g. Biber 1994) can be used to identify the first three types of informality (by counting & comparing over corpora the relative frequency of, for instance, 2nd person pronouns, vocatives, abbreviations, ellipses, taboo words and maledicta), the register-related source of informality cannot be effectively identified unless the alternate surface forms of 2nd person pronouns, vocatives, taboo words, etc. are taken into account. While current stylometric approaches, more particularly, correctly observe that 2nd person pronouns and maledicta occur much more frequently in IRC than in other written mediums, the proportion of, for instance, dialectical and vernacular variants of 2nd person references must be determined to identify register variation.

In order to tackle the latter type of variability more adequately, an operational measure of linguistic overlap was introduced in Geeraerts, Grondelaers & Speelman (1999), which builds on the notions of *onomasiological profile* - the set of denotationally equivalent designations of a concept/function and their respective relative frequencies in a corpus - and *uniformity*, i.e. the quantified overlap between onomasiological profiles. This methodology was subsequently used to compare lexical, morphological, syntactic & phonological preferences in Belgian IRC logs and other modes of written communication (viz. UseNet, regional & national popular newspapers, and quality newspapers). What these calculations reveal is that the four types of informality do not coincide in Belgian Dutch IRC: no matter how informal chat is situation-, medium-, and speakerwise, from the register point of view, chatters do not manifest colloquial or non-standard linguistic behavior to the degree one might anticipate. This leads us to the conclusion that the linguistic specificity of chat is not, or not in the first place, determined by register choices, but by the production demands (specifically speed, and turn-taking efficiency) of spoken conversation.

References
Biber, D. 1994. An analytical framework for register studies. In D. Biber & E. Finegan (eds.), Sociolinguistic perspectives on register, 31-56. OUP.
De Gryse, P. 2000. *Vloeken en verwensingen op IRC*. M.A. Thesis University of Leuven.
Geeraerts, D., S. Grondelaers & D. Speelman 1999. *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbaltermen*. Amsterdam: Meerteninstituut.
Hentschel, E. 1998. "Communication on IRC". Online publication on http://viadrina.euv-frankfurt-o.de/~wjournal/irc.html