Modeling register at the interface between social factors and grammar

Accounting for social variation at the level of syntax has been considered a challenge, both from variationist sociolinguists (see conceptual discussion in e. g. Lavandera 1978, Tagliamonte 2006) and from syntacticians (e. g. Newmeyer 2003). This challenge is particularly complex within the generative paradigm and its well-known idealization assumptions of grammatical competence. In this paper we take inspiration from classic variationist sociolinguistic work on style shifting (Labov 1966, Bell 1984), and propose a model in which we do not assume a direct link between social factors and syntax, but an indirect relation, where the effect of social factors on syntax is mediated via register, i. e. via the part these social factors play in the situational-functional setting with respect to register variation / style shifting. Register variation is here understood as intra-individual variation in linguistic behaviour which serves to address the communicative needs of the situation, resulting in socially recurring varieties or codes with particular co-occurrences of linguistic phenomena due to the situational-functional parameters (Biber and Conrad 2009; Lüdeling et al. 2022). Our proposed model does not need to resort to the assumption of different grammatical systems in order to account for social variation at the level of syntax but builds on the probabilistic approach of register variation (Biber 1988).

In order to assess to which extent empirical facts are in accord with such a model, we test whether we find (i) register-related differences, (ii) overall social differences, and (iii) socio-register interactions at the level of syntax. We are expecting register differences as well as socio-register interactions. Furthermore, we are comparing two different speech communities from two languages, German from Berlin and Persian from Tehran. We have selected clausal embedding complexity as a *metrical* dependent variable, counting whether and, if so, how many embedded clauses per utterance were used (Verhoeven & Lehmann 2018, Adli 2022, Lehmann 2024).

Our data stems from the German and Persian parts of the Lang*Reg corpus (Adli et al. 2023). 12 German participants and 20 Persian participants have been recorded in six well-defined situations: four spontaneous conversations, including with a professor in his office, a taxi driver during a ride, with a close friend, and with an unacquainted person from similar educational and age background, as well as one narrative monologue directed to a friend and one letter to a friend. We used gender match vs. mismatch between speaker and interlocutor as a social variable (Nanbaksh 2011), which is interactional by nature, reflecting the gender combination of both interlocutors. The mean speaker clausal complexity per situation is shown in Figure 1.

We fitted the results for clausal embedding complexity by a generalized linear mixed-effects model (Poisson distribution) using COMMUNICATIVE SITUATION – itself a multidimensional construct – and GENDER MATCH as fixed factor and NUMBER OF CLAUSAL EMBEDDING PER NON-EMBEDDED CLAUSE as the dependent variable with the random intercept of SPEAKER. The results show, in both languages, register effects for social hierarchy, viz. professor conversation vs. taxi driver conversation (German: $p<0.001$; Persian: $p<0.001$), as well as an effect of mode, viz. oral vs. written friend storytelling (German: $p<0.05$; Persian: $p<0.01$). Given the small sample size, we use an α level of 10 %. We find interaction effects between GENDER MATCH and specific registers in both languages: in Persian for taxi driver conversations ($p<0.07$) and in German for oral storytelling ($p<0.05$). The model with the factor GENDER MATCH has a better fit in Persian, and marginally also in German. In today's Tehran, speakers tend to use *more* clausal embedding in a dialogue situation with a taxi driver, typically perceived lower on the social hierarchy, if the interlocutors have the same

gender. In today's Berlin, a speaker uses *less* clausal embedding when having a monologue directed to a friend of the same gender. We will discuss the social-interactional meaning of gender match in each of these situations.

In conclusion, while speakers employ characteristic registers associated with given communicative situations, social factors may shift the probabilistic composition of registers in the multi-dimensional space of linguistic features (Biber 1988).

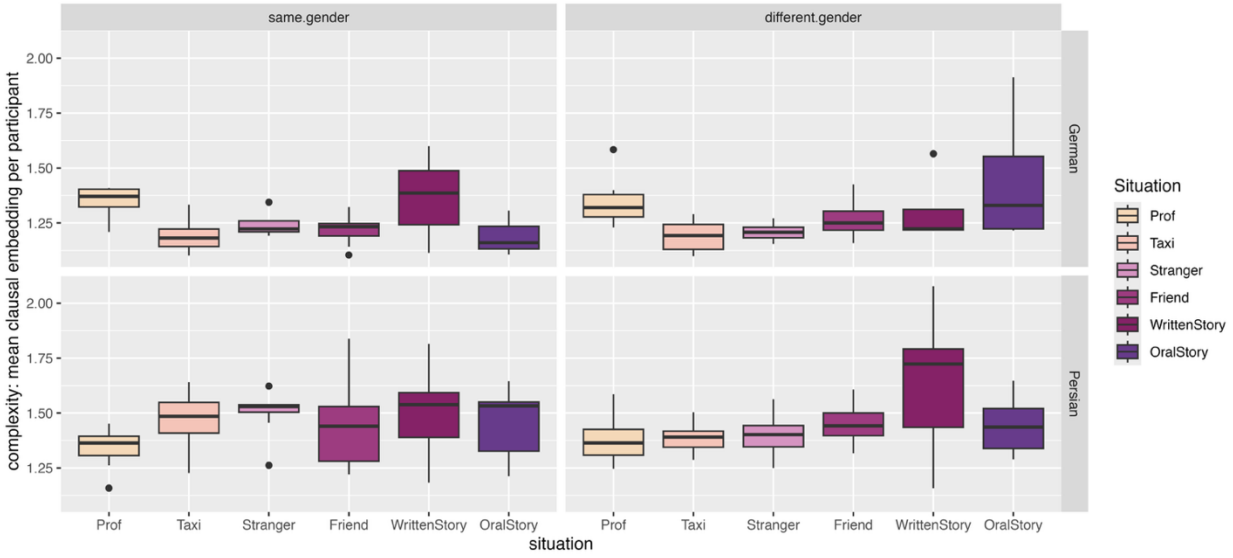

Figure 1: Mean clausal embedding per participant in each situation grouped by same or different gender

Adli, A., E. Verhoeven, N. Lehmann, V. Mortezapour, & J. Vander Klok. (eds.). 2024. Lang*Reg: A multi-lingual corpus of intra-speaker variation across situations. Version 0.2.0. [Data set]. Zenodo. Berlin: Humboldt-Universität zu Berlin. doi.org/10.5281/zenodo.13889198

Adli, A. (2022). Coherence and implicational hierarchies in the speech of the very old. In Beaman, K. V. & Guy, G. R. (eds.), *The Coherence of Linguistic Communities: Orderly Heterogeneity and Social Meaning.* New York & London: Routledge. 178-199.

Bell, A. (1984). Language style as audience design. *Language in Society* 13(2): 145–204.

Biber, D., 1988. Variation across speech and writing. Cambridge University Press, Cambridge.

Biber, D. & Conrad, S. (2009). *Register, Genre, and Style.* Cambridge: Cambridge University Press.

Labov, W. (1966). *The Social Stratification of English in New York City.* Washington DC: Center for Applied Linguistics.

Lavandera, B.R., 1978. Where does the sociolinguistic variable stop? Language in Society 7, 171–182.

Lehmann, N. 2024. The intricacies of register variation across languages. PhD thesis. Humboldt-Universität zu Berlin: Berlin.

Lüdeling, A., et al., 2022. Register: Language Users' Knowledge of Situational-Functional Variation: Frame text of the First Phase Proposal for the CRC 1412. REALIS 1, 1–59.

Nanbaksh, G. (2011). *Persian address pronouns and politeness in interaction*. PhD dissertation, University of Edinburgh.

Newmeyer, Frederick J. (2003). Grammar is grammar and usage is usage. *Language* 79(4): 682–707.

Tagliamonte, S.A., 2006. The linguistic variable, in: Analysing Sociolinguistic Variation, Key Topics in Sociolinguistics. Cambridge University Press, pp. 70–98.

Verhoeven, E., Lehmann, N., 2018. Self-embedding and complexity in oral registers. Glossa: a journal of general linguistics 3, 1–30.