HOW MUCH LANGUAGE USE IS ACTUALLY ON SOCIAL TOPICS?

ANNA SZALA[*1,2], MAREK PLACIŃSKI[1], ALEKSANDRA PONIEWIERSKA[1], ALEKSANDRA SZCZEPAŃSKA[1] and SŁAWOMIR WACEWICZ[1]

[*]Corresponding Author: szala@oakland.edu
[1]Centre of Language Evolution Studies, Nicolaus Copernicus University, Toruń, Poland
[2]Department of Psychology, Oakland University, Rochester, MI, USA

1. Introduction

How much language use is on social topics? Dunbar and colleagues (1997) claimed that "socially relevant topics" accounted for about two-thirds of time spent on conversation, and later studies largely confirmed this pattern, including cross-culturally (e.g., Kubik et al., 2018). In addition to its considerable popular impact, Dunbar et al.'s study broke new theoretical grounds, helping motivate the evolutionary Social Brain theory and the influential "gossip" theory of language origins (however, the term "gossip" is problematic and should be used with caution – see esp., Dores Cruz et al., 2021).

Here, we revisit Dunbar's question about the proportion of conversational time devoted to social topics. In this effort, we are motivated by the substantial limitations of both the dataset and approach of the original study. Dunbar et al. (1997) relied on a relatively small number of conversations (N = 45), collected exclusively in open public environments, between a sample of participants with a very limited demographic and geographic distribution. Secondly, their definition and operationalisation of "social topics" lacked clarity and sufficient connection to the proposed adaptive functions of language. For example, Dunbar et al.'s classification of the categories of sport/leisure, culture/art/music, or politics as non-social appears arbitrary in the light of our own data (see below) and grouping "personal experiences" within the larger category of social topics does not appear to demonstrate a straightforward link to the proposed adaptive social functions of language that motivate the evolutionary conclusions drawn from this study.

2. Materials and methods

We used Spokes (Pęzik, 2014), a corpus of Polish informal, casual conversations (N = 668; over 2.6 million word tokens) with speakers from a large variety of Polish social backgrounds. Mindful of the limitations of Dunbar's original approach, we re-defined "social topics" as conversations concerning 1) people that at least one participant knew in person (e.g., neighbour), 2) people they did not know in person (e.g., musician), and 3) social groups they either knew or did not know in person (e.g., classmates, football teams). As a consequential decision, content related to the speakers' personal experiences which did not involve others was considered non-social. We randomly selected 56 conversations (ca. 10% of the dataset excluding the training set), which were independently assessed by two judges, who rated each text line in each transcript as concerning a social or non-social topic.

3. Results

The expert judges converged on 85% of their judgments (moderate interrater agreement; $\kappa = 0.68$). Portions of conversations classified differently by the judges were removed from the analysis. The total number of word tokens was 52436 for social topics (50.9%) and 50528 for non-social topics (49.1%). We fitted age and gender to a logistic regression model to check how they link with conversing on social topics. Age positively correlated with conversing on social topics ($\beta = 0.015$, $p < 0.001$); gender did not show an effect but a trend was for males to speak less on social topics ($\beta = -0.073$, $p = 0.09$).

4. Conclusions

We found a roughly equal share of conversation devoted to social versus non-social topics, which differs considerably from the roughly 2/3 reported by Dunbar et al. (1997). One key difference was our classification of "personal experiences" (content limited to the individuals involved in the conversation) as non-social. Unsurprisingly, the proportion of social topics in conversation is very sensitive to the definition applied, and since developing a standard multipurpose definition does not appear feasible (cf. e.g. phatic communication is technically not on "social topics"), further research should, in particular, compare the different definitions of "social topics" that are directly rooted in the different socially adaptive functions of language.

Acknowledgements

References

Dores Cruz, T. D., Nieper, A. S., Testori, M., Martinescu, E., & Beersma, B. (2021). An Integrative Definition and Framework to Study Gossip. *Group & Organization Management, 46*(2), 252–285. DOI: 10.1177/1059601121992887

Dunbar, R. I. M., Marriott, A., & Duncan, N. D. C. (1997). Human conversational behavior. *Human Nature, 8*, 231–246. DOI: 10.1007/BF02912493

Kubik, B. G., Pápay, B. T., Takács, K., & Galántai, J. (2018). The Gossip Hypothesis. [Prezi slides]. European Research Council. www.prezi.com/bs2dnr14_j2x/the-gossip-hypothesis.

Pęzik, P. (2014, October 24–25). Spokes – a search and exploration service for conversational corpus data. [Selected Papers]. CLARIN 2014: Soesterberg, The Netherlands. www.ep.liu.se/ecp/116/009/ecp15116009.pdf.