Exploring the content of casual Polish conversations

Szala, A.[*1], Wacewicz, S.[1], Placiński, M.[1], Poniewierska, A. E.[1], Schmeichel, A.[1], Stefańczyk, M.[2], & Żywiczyński, P.[1]

[*] Corresponding Author: szala@oakland.edu
[1] Centre of Language Evolution Studies, Nicolaus Copernicus University, Toruń, Poland
[2] Institute of Psychology, University of Wrocław, Poland

In a seminal study, Dunbar, Marriott, and Duncan (1997) found that ≈67% of conversational time is spent on discussing social topics, an estimate that exerted a significant influence on theories of the evolution of the human brain, cognition, and language. However, the work by Dunbar et al. (1997) had substantial limitations (small, demographically and geographically limited sample; data collected exclusively in open environments; unclear operationalisation of "social topics"), which motivated our recent preliminary study on this topic (Szala et al. 2022). Here, we report a follow-up, full-scale study revisiting this issue.

Similarly to Szala et al. (2022), we used Spokes (Pęzik, 2012; 2014). Spokes is a corpus of 669 Polish informal conversations based on live recordings of casual speech obtained in private as well as public places, with speakers from a variety of Polish demographic backgrounds, including age (ranging from 1 to 99 years), and education levels spanning from none to higher education. Some conversations in Spokes were recorded surreptitiously, with consent and demographic data provided after the recording. Spokes is manually divided into lines so as to mark alternating contributions of individual speakers.

In our study, we excluded any conversations too short or too long to reliably code for conversation topics, operationalised as conversations comprising 50 or fewer lines, or containing at least one line with more than 150 word tokens. In the resulting dataset of 535 Polish conversations, every line was coded by two native Polish-speaking coders to ensure reliability. Our main distinction was between social vs non-social topics, understood in terms of information content. Social topics were operationalised as "sharing information related to self and other people", as opposed to sharing other information (e.g.,

factual conversation about technology). We excluded all lines of text with rater disagreement, i.e. coded as social by one rater but non-social by the other. Two coders agreed on 71% of lines in the dataset. This resulted in a dataset comprising 197,621 lines of text, with a mean conversation length of ca. 367 lines, and a mean line length of ca. 9 word tokens. Our study was pre-registered (Szala et al., 2023; https://osf.io/kjf4e), and the dataset and coding scheme were made publicly available (https://osf.io/mqs5k/).

In contrast to Dunbar et al. (1997), who found that ≈67% of conversation time is spent on social topics, our study indicates that social topics can account for as much as 85% of conversation. An important aspect of our work is how "social topics" are operationalized. In Szala et al. (2022), we *excluded talking about oneself, i.e. counted it as non–social*, and found that 51% of conversations were devoted to social topics so defined. However, since self-disclosure is pivotal in forming social relationships and in particular plays a vital role in reputation building, here we decided to include this category under the rubric "social". This further underscores the important point that the proportion of "social" to "non-social" topics in conversation is highly sensitive to how "social topics" are defined. In this study, in line with our evolutionarily motivated research question, we defined "social topics" focusing on subjects related to social bonding, cooperation, and human interactions, which may not be universally applicable to all theoretical perspectives on language use.

Our main result (85% of conversation devoted to social topics) confirms, and actually exceeds, Dunbar et al.'s original estimate of 67%, and aligns with other studies examining social discourse, such as 76% in Dahmardeh & Dunbar (2017), reinforcing the general conclusion that a majority of topics in casual conversations tend to be of a social nature. The consistent findings across a range of studies that use varied datasets and methodologies underscore the crucial role that exchanging social information plays in human communication across different contexts and populations. This in turn lends indirect support to theories that stress the role of social ecology in shaping hominin cognitive evolution. We complement our study with exploratory analyses that include demographic factors (gender, age, and education) and a finer subcategorisation of social topics into information on individuals (i) participating in the conversation, (ii) not participating in the conversation but known to the conversants, (iii) not known to the conversants (e.g. celebrities, fictional characters).

Acknowledgements
This research was supported by the Polish National Science Centre under grant agreement UMO-2019/34/E/HS2/00248.

References

Dahmardeh, M. & Dunbar, R. I. M. (2017). What shall we talk about in Farsi? *Human Nature, 28*(4), 423–433. https://doi.org/10.1007/s12110-017-9300-4

Dunbar, R. I. M., Marriott, A., & Duncan, N. D. C. (1997). Human conversational behavior. *Human Nature, 8*(3), 231–246. https://doi.org/10.1007/BF02912493

Pęzik, P. (2012). Jezyk mówiony w NKJP [Spoken language in National Corpus of Polish Language]. In: A. Przepiórkowski, M. Bańko, R. Górski, & B. Lewandowska-Tomaszczyk (Eds.). *Narodowy Korpus Jezyka Polskiego* [National Corpus of Polish Language], 37–47. Wydawnictwo Naukowe PWN, Warszawa.

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

Szala, A., Placiński, M., Poniewierska, A., Szczepańska, A., & Wacewicz, S. (2022). How much language use is actually on social topics? In: A. Ravignani, R. Asano, D. Valente, F. Ferretti, S. Hartmann, M. Hayashi, … S. Wacewicz, (Eds.). *The Evolution of Language: Proceedings of the Joint Conference on Language Evolution (JCoLE)* (705–707). Joint Conference on Language Evolution (JCoLE). https://doi.org/10.17617/2.3398549

Szala, A., Placiński, M., Żywiczyński, P., Poniewierska, A. E., Schmeichel, A., & Wacewicz, S. (2023, January 25). *How much language use is actually on social topics: Human conversational behavior revisited.* https://doi.org/10.17605/OSF.IO/KJF4E