PSE v1.0: The first open access corpus of public service encounters

Ingrid Espinoza[1], Steffen Frenzel[1], Laurin Friedrich[1], Wassiliki Siskou[1,2],
Steffen Eckhard[3], Annette Hautli-Janisz[2]

[1]University of Konstanz   [2]University of Passau   [3]Zeppelin University

firstname.lastname@{uni-konstanz.de|uni-passau.de|zu.de}

Abstract

Face-to-face interactions between representatives of the state and citizens are a key intercept in public service delivery, for instance when providing social benefits to vulnerable groups. Despite the relevance of these encounters for the individual, but also for society at large, there is a significant research gap in the systematic empirical study of the communication taking place. This is mainly due to the high institutional and data protection barriers for collecting data in a very sensitive and private setting in which citizens request support from the state. In this paper, we describe the procedure of compiling the first open access dataset of transcribed recordings of so-called Public Service Encounters in Germany, i.e., meetings between state officials and citizens in which there is direct communication in order to allocate state services. This dataset sets a new research directive in the social sciences, because it allows the community to open up the black box of direct state-citizen interaction. With data of this kind it becomes possible to directly and systematically investigate bias, bureaucratic discrimination and other power-driven dynamics in the actual communication and ideally propose guidelines as to alleviate these issues.

Keywords: Public Service Encounters, Transcripts, Computational Social Science

1. Introduction

Public Service Encounters (PSEs) are referred to as the "ground floor of government" (Eckhard et al., 2022, p. 3): citizens directly interact with a representative of the state in order to get support in personal matters, for instance when applying for social welfare. The representatives are responsible for applying laws and decrees to the personal situations of individuals (Bruhn and Ekström, 2017, p. 211) and are bounded to advocate for social equity (Raaphorst, 2021; Sumra, 2019; Zacka, 2019, p. 454). This makes PSEs one of the central hinges where governments attempt to grant citizens the support they need in order to be equal members of society.

Government officials working at the front line of public service delivery are therefore at the center of the state's actions in policy implementation, because they decide on the concrete application of services to individual cases for which policies can only provide abstract and vague guidelines (Hupe and Hill, 2007; Lipsky, 2010; Maynard-Moody and Musheno, 2000). For doing so, the officials have a certain leeway in the decision-making process, which is necessary to allocate the available resources appropriately among citizens. One of the challenges is that public officials are, as all human beings, subject to unconscious biases, which potentially show up in the face-to-face interactions (Hardin and Banaji, 2013). These biases can lead to an uneven distribution of access to public services to the disadvantage of marginalized and minoritized social groups, a fact that has been established mainly with experimental work in social science (Hemker and Rink, 2017; Thomann and Rapp, 2018; Adam et al., 2021). However, these experiments leave out the central intercept in the process, namely the communicative interaction between officials and citizens. How public servants communicate, that is whether they use more comprehensive or relational communication, has been shown to affect citizen satisfaction (Eckhard and Friedrich, 2022). It is therefore crucial to shed light on the causes of inequality, with a systematic study of the patterns of communication in PSEs and the ways in which interactional power dynamics are constructed and expressed in language.

This paper contributes to this significant research gap by providing the first ever publicly released dataset on PSEs containing transcribed recordings of 106 meetings across public service offices in four cities in Germany. This dataset is particularly significant given the high institutional barriers for collecting this kind of highly sensitive data (both for government officials and citizens) and the tight ethical and data protection restrictions that need to be met for eliciting and processing it. In the present paper we present the steps for eliciting this data and the core characteristics of the dataset.

2. Related work

In social science there is a significant amount of work on the factors that have an impact on public service delivery. For instance, studies have shown that putative members of historically disadvan-

13315

(1)   a.   [00:03:46 - 00:03:57]
           Official: *Well, after it became clear that your daughter should live with you, you should have applied for custody quickly.*
      b.   [00:04:00 - 00:04:14]
           Citizen: *Uh, I tried, at the youth welfare office, but the youth welfare office said: "Ah yes, this all has to go before the family court and then this has to be clarified and that has to be clarified".*
      c.   [00:04:15 - 00:04:16]
           Official: *What is "this and that"? What do you mean by that?*
      d.   [00:04:16 - 00:04:19]
           Citizen: *How I feel about the child and everything.*
      e.   [00:04:20 - 00:04:24]
           Official: *How do you feel about the child?*
      f.   [00:04:26 - 00:04:30]
           Citizen: *Actually quite good, never – never had any problems or anything.*

taged groups (across countries and settings) are less likely to receive a response from state officials (Distelhorst and Hou, 2014; Einstein and Glick, 2017; Hemker and Rink, 2017). Gender, ethnicity and language fluency have also been established as factors contributing to bureaucratic discrimination (Adam et al., 2021; Fernández-i Marín et al., 2021; Jilke and Tummers, 2018), as well as administrative processes that disadvantage vulnerable groups (Bell and Smith, 2022; Thomann and Rapp, 2018). A small number of qualitative ethnographic studies synthesize insights on these encounters in a subjective, interpretative way (Epp et al., 2014; Maynard-Moody and Musheno, 2000; Nisar and Masood, 2020; Nisar, 2020; Perelmiter, 2022).

More recently, the linguistically-driven investigation of communication between representatives of the state and citizens has gained some attention: Voigt et al. (2017) and Rho et al. (2023) focus on police cam footage to investigate whether black citizens are treated less respectfully, but the data is only available as n-grams. Siskou et al. (2022) present a plain language score to measure comprehensibility during PSEs, the data on which we build the current paper on. In terms of an empirically driven theory building, Eckhard et al. (2022) propose a taxonomy for describing PSE language, focused on the perceptions of citizens.

Taken together, this paper fills a research gap between social science, corpus linguistics, and Natural Language Processing: so far, systematic studies into the micro-level of public service delivery, namely PSEs, were impossible to conduct because data was lacking. Even if data was available, it was not published in a way that would allow for an analysis of the dynamics of the conversation. With PSE v1.0 we move a significant step ahead in that we deliver the first open access resource that allows for an in-depth analysis of the communication, not only in terms of n-grams, but full transcripts of actual recordings. We thus pave the way for new research in NLP and social science that will allow us to move beyond experiments to shed light on the ways that government officials engage with citizens.

3.   The PSE Corpus, v1.0

3.1.   The genre

Public service encounters are generally in-person meetings and have been scheduled in advance, usually initiated by a citizen's application for social benefits and a subsequent interview by the public official to determine eligibility and extent of the support. Example (1)[1] illustrates an exchange between a citizen and a public service official in which the custody of the citizens daughter is discussed (the German original is in Example (3) in the Appendix). Typical conversation patterns are the formulation of obligations by the official in (1a) and information seeking moves by the official in (1c) and (1e), which the citizen is expected to respond to. This conversational structure shows a clear hierarchy between official and citizen.

In contrast to these highly hierarchical dynamics, there are also more informal exchanges, for instance shown in Example (2)[2]: Citizen and official have a light conversation about family life and upcoming festivities, which is not driven by information seeking moves from the official, but is meant to establish a personal bond between the interlocutors and to create a relaxed conversational atmosphere. This type of conversation is typical, in particular for the beginnings and ends of PSEs.

As can be seen here, in particular in Example (1), the content of these conversations is highly personal. As a consequence, there is an enormous

---

[1] Translated to English from German original with transcript id '202111091430sxj2ulsjMAYTS'.

[2] Translated to English from German original with transcript id '202111240815el4d0y4nMAYMS'.

(2)   a.   [00:00:05 - 00:00:07]
            Citizen: *Yes, in four weeks!*
      b.   [00:00:07 - 00:00:10]
            Official: *Crazy, completely crazy!*
      c.   [00:00:10 - 00:00:14]
            Citizen: *[laughs] And the children are already going crazy at home. I mean it's not normal anymore!*
      d.   [00:00:16 - 00:00:19]
            Official: *Already? Because of Christmas?*
      e.   [00:00:19 - 00:00:21]
            Citizen: *Yes, well I have decorated the house already, you know? So yes, they are really exited.*
      f.   [00:00:22 - 00:00:23]
            Official: *Ah, nice!*

amount of work involved in (a) getting public offices to collaborate, (b) get citizens and public officials to agree to have their meetings recorded and (c) ensure rigid data anonymisation applying to tight data protection regulations.

3.2. Data collection

The data collection for PSE v1.0 was conducted between 2021 and 2023. The necessary steps are elaborated on in the following.

Gaining access to institutions For recruitment, we contact government agencies all over Germany, explaining the background and goals of the project. Bringing them on board entails a myriad of different steps that can consume months of preparatory work, such as multiple discussion rounds with different veto players within the organization that can impede an institution from providing access to this encounters between civil servants and citizens, e.g., CEOs, staff councils and the concerning heads of department. In case of a positive review, employees are informed about the project in a number of (on-site) information events.

Recruiting officials and citizens The process of recruiting officials and citizens is self-selecting – at no point in the procedure do we or heads of departments and officials have the right or leverage to request participation. This means that we cannot guarantee the absence of a self-selection bias in the data collection. We do have requirements on the number of employees per institution that need to take part: a minimum of five employees per agency needs to sign up in order to avoid skewed analyses and to maintain data protection and anonymity.
Once the agency is on board, the officials receive extensive briefing on the data protection regulations and the technical handling of the recordings. The citizens are recruited by the public service officials taking part in the study in the scheduled meetings they have with them. The citizens are thoroughly informed about the use of the data. They are also explicitly informed that the recording of a conversation can be stopped at any time. All participants (officials and citizens) contribute voluntarily and there is no incentive involved in having a meeting recorded, monetary or otherwise. In addition, we distribute a paper-based survey to collect additional data (more information below). Recordings only start after consent of both parties.

Recording For recording we use dictaphones that are specifically prepared by project team members in order to adhere to the data protection regulations: All devices are prepared so that no form of manipulation can be conducted by the employees. They are only allowed to turn the device on and off and to make recordings but there is no possibility to play back or delete recordings.

Data transfer After the end of the first phase, the audio devices are collected by members of the project and sealed. During transport, the devices have to be stored securely so as to avoid loss or theft. Once at university, the data is transferred to a local machine in a secure data room with access restricted to project members.

Transcription The transfer to and from the transcription company is encrypted and the company adheres to the strict GDPR regulation established for the project. The transcription follows the guidelines of Dresing and Pehl (2015) and captures the verbatim content of the audio file (with grammatical errors, filler words, stutter and repetitions), without any modifications or omissions. As long as the transcripts are not anonymised they have to be stored in the secure data room.

Anonymisation Once transcribed, the data is anonymised by project members, following a detailed set of instructions on when and how to replace sensitive content. This applies not only to actual names of the speakers, but also to locations, dates and other information like names of employers that would allow outsiders to trace back to individual people. They are replaced with tags

that represent the original data, e.g., [Name1], [Location1], etc. This anonymisation process is done manually and all edits are double-checked by at least one other project member before the data is stored outside the secure data room.

3.3. Data cleaning and release

Once the data is anonymised, it is transferred to regular university servers. We perform final steps of data cleaning, e.g., special characters are removed and irregularities in the transcription are corrected. Contributions of rarely used interpreters in our corpus are clearly marked with a distinct speaker tag in the transcripts. While utterances in German are fully transcribed, those in any other language are consistently denoted by the placeholder <foreign language> in the content of the utterance. The corpus is available both as plain text transcripts and in an XML format that contains additional information. We manually insert a gender tag for each speaker, based on the speaker tags provided in the transcription. This tag is necessary to be able to make gender-specific comparisons later on. Alongside the original German transcripts, the corpus also includes ethics clearance and the project-related data protection regulations. The dataset is available under the CC0 1.0 Universal license at Harvard Dataverse.[3] In accordance with the GDPR restrictions, we delete the audio data after completed data cleaning and segmentation. But we also need to make sure that study participants can at any point in the future withdraw their data from the corpus. The information which data point corresponds to which participant is stored safely in the secure data room.

3.4. Survey data

As part of the study, both officials and citizens are asked to complete a survey. The survey allows us to collect additional information from the officials such as as job satisfaction, emotional and physical exhaustion, and the level of work load. We also collect information on the officials' level of client aversion and emphatic compassion. With regard to the citizens, we elicit demographic information (gender, age cohort, migration background), as well as information on perceptions of satisfaction with the encounter, fairness of the communication, and the official's responsiveness. We also ask for information of how much previous contact they had with the agency, in addition to factors related to more general attitudes such as self-efficacy and trust in government institutions. This information paves the way for future studies that bring together information on the dynamics of an interaction, the official's background and the judgement of the client regarding the communication. The
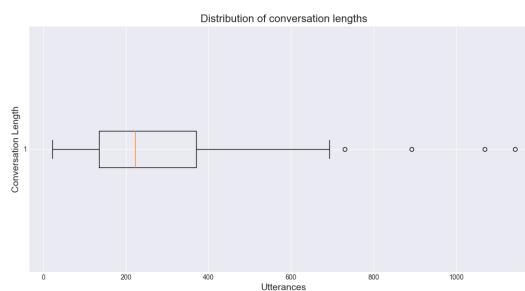
Figure 1: Distribution of conversation length in PSE v1.0

survey data will also be made available via open access, but in a later version of the dataset.

3.5. Statistics

The dataset consists of a total of 106 conversations with 433,780 tokens in around 31,451 turns. The average sentence length is 5.4 tokens; the average turn length is 2.5 sentences. There are a total of 159 different speakers in the dataset, of which 133 are clients and 26 are employees from the four public agencies that participated in the project. Occasionally, interpreters were used because the citizens did not speak German. The interviews were recorded at the following German authorities: Jobcenter Mayen-Koblenz (41 conversations), Jobcenter Stuttgart (24 conversations), Bürgerbüro Konstanz (19 conversations) and Stadtamt Durlach (22 conversations). Other institutions that will further diversify the dataset are still in the data collection phase and will be published in v2.0 of the dataset.

In total, there are around 11,186 turns from male speakers (35,6%) and around 20,133 turns from female speakers (64%). The distribution of speaker turns between citizens and officials is very balanced, with around 15,481 (49,2%) turns from officials and around 15,891 (50,5%) turns from citizens. Overall, the average conversation length is 296 turns, with a median of 236 turns and the IQR ranging from 126 to 393 turns. As can be seen in Figure 1, the dataset features two outliers with more than 1,000 turns, the longest conversation contains 1,143 turns.

The largest share of data comes from job centres (approx. 61%) ('Jobcenter') which in Germany are responsible for granting benefits to people without employment and trying to (re)integrate them into the labour market. Important is here that citizens have regular, scheduled appointments and might have been in contact with 'their' official over a longer period of time.

4. Avenues for future work

In addition to increasing the corpus size, there are multiple avenues for future work: For instance, Eckhard et al. (2022) propose a distinction between the *informational* and the *relational component* of PSEs, with the former focusing on the question of how efficiently and comprehensibly information is exchanged, and the latter focusing on the negotiation of hierarchical relationships and personal bonds between officials and citizens. Operationalising these dimensions is one of the key areas where social science and language technology can come together. One avenue is to measure the impact of small talk as a means of personal bonding on the overall satisfaction of the citizens. One can also explore the disparities in respectful treatment to the disadvantage of minoritized social groups in PSEs. Another avenue is to develop technology that runs in real time as the conversation unfolds and shows participants potentially biased or discriminatory language, potentially having a positive impact on how public service encounters are conducted.

Acknowledgements

5. Bibliographical References

Christian Adam, Xavier Fernández-i Marín, Oliver James, Anita Manatschal, Carolin Rapp, and Eva Thomann. 2021. Differential discrimination against mobile EU citizens: experimental evidence from bureaucratic choice settings. *Journal of European Public Policy*, 28(5):742–760.

Elizabeth Bell and Kylie Smith. 2022. Working within a System of Administrative Burden: How Street-Level Bureaucrats' Role Perceptions Shape Access to the Promise of Higher Education. *Administration & Society*, 54(2):167–211.

Anders Bruhn and Mats Ekström. 2017. Towards a Multi-Level Approach on Frontline Interactions in the Public Sector: Institutional Transformations and the Dynamics of Real-time Interactions. *Social and Policy Administration*, 51(1).

Greg Distelhorst and Yue Hou. 2014. Ingroup Bias in Official Behavior: A National Field Experiment in China. *Quarterly Journal of Political Science*, 9(2):203–230.

T. Dresing and T. Pehl. 2015. *Praxisbuch Interview, Transkription & Analyse: Anleitungen und Regelsysteme für qualitativ Forschende*. Dresing.

Steffen Eckhard and Laurin Friedrich. 2022. Linguistic Features of Public Service Encounters: How Spoken Administrative Language Affects Citizen Satisfaction. *Journal of Public Administration Research and Theory*.

Steffen Eckhard, Laurin Friedrich, Annette Hautli-Janisz, Vanessa Mueden, and Ingrid Espinoza. 2022. A Taxonomy of Administrative Language in Public Service Encounters. *International Public Management Journal*.

Katherine Levine Einstein and David M Glick. 2017. Does Race Affect Access to Government Services? An Experiment Exploring Street-Level Bureaucrats and Access to Public Housing. *American Journal of Political Science*, 61(1):100–116.

Charles R Epp, Steven Maynard-Moody, and Donald Haider-Markel. 2014. *Pulled Over: How Police Stops Define Race and Citizenship*. University of Chicago Press.

Xavier Fernández-i Marín, Christoph Knill, and Yves Steinebach. 2021. Studying Policy Design Quality in Comparative Perspective. *American Political Science Review*, 115(3):931–947.

Curtis D. Hardin and Mazarin R. Banaji. 2013. The Nature of Implicit Prejudice: Implications for Personal and Public Policy. In *The behavioral foundations of public policy*, pages 13–31. Princeton University Press.

Johannes Hemker and Anselm Rink. 2017. Multiple Dimensions of Bureaucratic Discrimination: Evidence from German Welfare Offices. *American Journal of Political Science*, 61(4):786–803.

Peter Hupe and Michael Hill. 2007. Street-Level Bureaucracy and Public Accountability. *Public Administration*, 85(2).

Sebastian Jilke and Lars Tummers. 2018. Which Clients Are Deserving of Help? A Theoretical Model and Experimental Test. *Journal of Public Administration Research and Theory*, 28(2):226–238.

(3) a. [00:03:46 - 00:03:57]
        Mitarbeiter: *Gut, aber Sie hätten / man hätte ja eigentlich relativ zügig dann nach, ähm, als klar war, dass, äh, ihre Tochter zu Ihnen kann oder soll, das Sorgerecht beantragen können.*
     b. [00:04:00 - 00:04:14]
        Kunde: *Äh, habe ich probiert, äh, auf dem Jugendamt, aber das Jugendamt hat dann angefangen: "Ah ja, das muss alles vors Familiengericht und dann muss das geklärt werden und dann muss das geklärt werden".*
     c. [00:04:15 - 00:04:16]
        Mitarbeiter: *Was ist jetzt "das und das"? Was meinen Sie damit?*
     d. [00:04:16 - 00:04:19]
        Kunde: *Äh, wie ich zu dem Kind stehe und allem drum und dran.*
     e. [00:04:20 - 00:04:24]
        Mitarbeiter: *Wie stehen Sie denn zu dem Kind?*
     f. [00:04:26 - 00:04:30]
        Kunde: *Ei- eigentlich ganz gut, nie- nie Probleme gehabt oder so.*

Michael Lipsky. 2010. Street-Level Bureaucracy: Dilemmas of the Individual in Public Service. *American Political Science Association*, 76(1).

Steven Maynard-Moody and Michael Musheno. 2000. State Agent or Citizen Agent: Two Narratives of Discretion. *Journal of Public Administration Research and Theory: J-PART*, 10(2).

Muhammad A. Nisar. 2020. Phenomenology of the Stop: Street-Level Bureaucracy and Everyday Citizenship of Marginalized Groups. *International Review of Administrative Sciences*, 86(2).

Muhammad Azfar Nisar and Ayesha Masood. 2020. Dealing With Disgust: Street-Level Bureaucrats as Agents of Kafkaesque Bureaucracy. *Organization*, 27(6):882–899.

Luisina Perelmiter. 2022. "Fairness" in an Unequal Society: Welfare Workers, Labor Inspectors and the Embedded Moralities of Street-Level Bureaucracy in Argentina. *Public Administration and Development*, 42(1):85–94.

Nadine Raaphorst. 2021. Administrative Justice in Street-Level Decision Making: Equal Treatment and Responsiveness. In *The Oxford Handbook of Administrative Justice*, pages 1–30. Oxford University Press.

Eugenia H. Rho, Maggie Harrington, Yuyang Zhongand, and Jennifer L. Eberhardt. 2023. Escalated Police Stops of Black Men are Linguistically and Psychologically Distinct in their Earliest Moments. *Proceedings of the National Academy of Sciences of the United States of America*, 120(23).

Wassiliki Siskou, Laurin Friedrich, Steffen Eckhard, Ingrid Espinoza, and Annette Hautli-Janisz. 2022. Measuring Plain Language in Public Service Encounters. In *Proceedings of the 2nd Workshop on Computational Linguistics for Political Text Analysis (CPSS-2022) Potsdam, Germany.*

Kalsoom Sumra. 2019. Social Equity in Public Administration: Fairness, Justice and Equity, tools for social change. *Pakistan Administrative Review*, 3(1):1–15.

Eva Thomann and Carolin Rapp. 2018. Who Deserves Solidarity? Unequal Treatment of Immigrants in Swiss Welfare Policy Delivery. *Policy Studies Journal*, 46(3):531–552.

Rob Voigt, Nicholas P. Camp, Vinodkumar Prabhakaran, and William Hamilton. 2017. Language from Police Body Camera Footage Shows Racial Disparities in Officer Respect. *Proceedings of the National Academy of Sciences of the United States of America*, 114(25).

Bernardo Zacka. 2019. Street-Level Bureaucracy and Democratic Theory. In *Research Handbook on Street-Level Bureaucracy*, pages 448–461. Edward Elgar Publishing.

A. Appendix

Example (3) provides the German original of Example (1).