Contextualizing Smartphone-Typed Language With User Input Intention

Florian Bemmann
LMU Munich
Munich, Germany
florian.bemmann@ifi.lmu.de

Timo K. Koch
Institute of Behavioral Science & Technology
University of St. Gallen
St. Gallen, Switzerland
timo.koch@unisg.ch

Maximilian Bergmann
Institute of Behavioral Science & Technology
University of St. Gallen
St. Gallen, Switzerland
maximilian.bergmann@unisg.ch

Clemens Stachl
Institute of Behavioral Science & Technology
University of St. Gallen
St. Gallen, Switzerland
clemens.stachl@unisg.ch

Daniel Buschek
Department of Computer Science
University of Bayreuth
Germany
daniel.buschek@uni-bayreuth.de

Ramona Schoedel
University of Psychology Charlotte Fresenius University
Germany
ramona.schoedel@charlotte-fresenius-uni.de

Sven Mayer
TU Dortmund University
Dortmund, Germany
info@sven-mayer.com

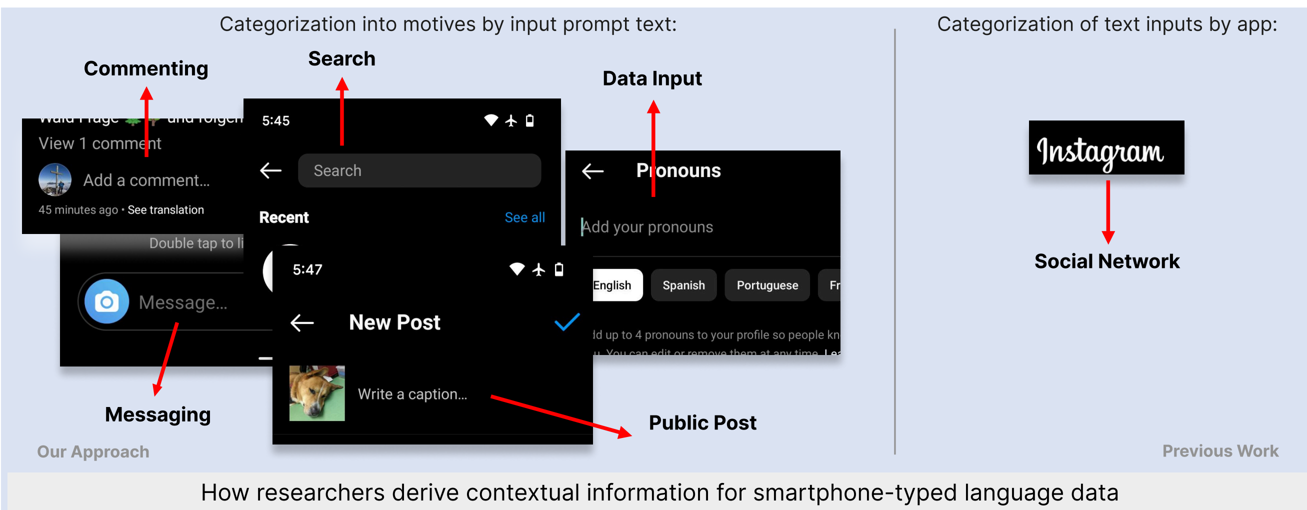

Figure 1: Instead of categorizing collected in-the-wild text input data by the originating app, we propose to regard text field's input prompt text. This figure shows on the example of the app Instagram, that text inputs into Instagram are not just social media contents such as posts and comments, but can also have other intentions such as messaging, search, and data input.

ABSTRACT

While the study of smartphone-typed language offers valuable insights for HCI and interdisciplinary research, existing data collection methods often fall short in providing context. Contextualizing text contents solely by the input target app does not allow for selecting them by the actual input intention of the user (e.g., sending a direct message, publishing a post). We present a method to study smartphone-typed language in the wild, which extracts the user's input intention from a text field's prompt text. Our approach enables distinguishing language contents by their channel (i.e., comments, messaging, search inputs), which allows filtering and pre-selecting text contents by the user's input intention and type of language. We provide software libraries and the underlying categorization schema of our method. Researchers can apply this on-device in studies relying on smartphone-typed language. We outline which further opportunities for adaptation our general procedure motivates, to facilitate interdisciplinary studies in the wild.

MuC '25, August 31-September 03, 2025, Chemnitz, Germany

ACM ISBN 979-8-4007-1582-2/25/08.
https://doi.org/10.1145/3743049.3748537

CCS CONCEPTS

• **Human-centered computing → Field studies**; **Text input**; **Smartphones**.

ACM Reference Format:
Florian Bemmann, Timo K. Koch, Maximilian Bergmann, Clemens Stachl, Daniel Buschek, Ramona Schoedel, and Sven Mayer. 2025. Contextualizing Smartphone-Typed Language With User Input Intention. In *Mensch und Computer 2025 (MuC '25), August 31-September 03, 2025, Chemnitz, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3743049.3748537

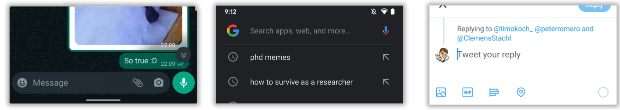

Figure 2: Screenshots of three text fields of the three Android apps: WhatsApp (left), Google Search (middle), and Twitter (right). All three text fields have input prompt texts, that give the user a hint about what the text field is intended to be used for.

1 INTRODUCTION AND BACKGROUND

Researchers from various fields are interested in studying our everyday language. Accordingly, digital footprints such as Facebook posts [28] or WhatsApp instant messages [13, 30] have been shown to provide useful information on personality traits or mental health [8, 28]. Various paths of research rely on smartphone-typed data to understand human behavior (e.g., [14]) and tackle issues such as mental health (e.g., [19]). Consequently, designing personalized interfaces with a great user experience in technical systems requires HCI and neighboring fields such as psychology or behavioral science, a profound understanding of users' psychological traits and states, mental models, and attitudes. The general idea behind most methodological approaches is to quantify raw language data [5, 7]. For example, in psychological research, typed language is mapped against dictionaries such as the Linguistic Inquiry and Word Count (LIWC) [22], as the basis for statistical analyses.

It is important to consider users' language *in context* as the meaning of textual contents might highly vary with the situation in which it is expressed. Also, specific research questions often require language of only a specific user intention. A traditional approach to analyzing language in context, often used for texts produced with smartphones, is to consider the app used for typing as contextual information. That is, each app is usually assigned to one meaningful category, such as Communication, Social Media, or Browser [27, 29]. However, this approach has a major disadvantage: Many apps can be used for different purposes, and therefore, the assignment to a single category as context information is a strong simplification. For example, the popular app Instagram is used frequently to post public content *and* to send direct messages. However, with the traditional app categorization approach, Instagram, as context for language, would be labeled either as communication *or* as social media. Furthermore, selecting content through specific platforms comes with systematic information gaps and thus biases (e.g., [9, 11]).

To avoid this simplification and thus inaccuracy in the analysis of smartphone-typed language, we propose a new approach that filters for *WHAT kind of content* users type instead of *WHERE the content was typed*. In this paper, we show how input prompt UI metadata can be used to contextualize language data. This new approach allows us to distinguish language contents by users' input intentions (i.e., publishing a post, commenting, direct messaging, performing a search). We specifically regard the input prompt text, i.e., the text that is visible as a background or placeholder inside a text field, as long as the user has not typed any content. We show some examples of these prompt texts in Figure 2. We have developed a prompt text categorization based on data from a large-scale field study, and we publish this together with an Android module library to retrieve smartphone text inputs and map them to users' intentions directly on the device. With this context-enriched keyboard logging and analysis method, we show the untapped potential for research in sensing language data. Carefully collected and high quality data is not only important for interdisciplinary research, but also essential to make the most of machine learning approaches and large language models.

Collecting comprehensive mobile language datasets is a challenging task. Current language analysis methods often rely on data donation and are thereby limited to specific platforms. To yield a platform-agnostic dataset that contains a holistic image of a user's typing language, smartphone APIs and research apps are more appropriate. The literature distinguishes three approaches to collecting text data via smartphones. Researchers can make users install a custom keyboard[1, 2], and thereby intercept the typed keys and log the users' keystrokes [3, 18, 25, 31]. This rather low-level data is rich and detailed: Researchers can comprehend the writing process in every detail, and the keyboard can also listen to other kinds of events related to text input like the usage of word suggestions and auto-completion events [3]. Alternatively, device APIs [20], such as the Android accessibility service API[3], offer application developers similar data without the need to exchange the installed keyboard [12, 17]. Applications can subscribe to text input events and be notified about each typed character. Past work also studies alternative approaches such as taking screenshots at a 5-second interval (e.g., [6]). While all these methods offer some way to access typing metadata, it has not yet been used to provide context to smartphone-typed language for interdisciplinary research. Studied application scenarios only encompass very specific use cases, such as UI automation (e.g., [23, 32]) and privacy-enhancing tools (e.g., [2, 10]).

2 CATEGORIZING INPUT PROMPT TEXTS TO CONTEXTUALIZE MOBILE TEXT INPUTS

In this paper, we propose an approach to contextualize mobile text inputs by the user's input intention. We therefore evaluate field hint texts, i.e., the placeholders displayed in smartphone text fields, to infer what kind of content a user typed. Our method is based

[1] https://developer.apple.com/documentation/uikit/keyboards_and_input/creating_a_custom_keyboard
[2] https://developer.android.com/guide/topics/text/creating-input-method
[3] https://developer.android.com/reference/android/accessibilityservice/AccessibilityService

on an input prompt text categorization scheme that we developed based on data from a large-scale mobile sensing field study. We incorporate our approach into a privacy-friendly mobile language logging system, where text is preprocessed on-device, and only word categories, but no raw texts, are sent to the researchers (c.f. [3]).

In more detail, our study provides (a) logging libraries for the Android operating systems to track language input data, and (b) a categorization approach for UI metadata to extract information on the user's motivation to produce text based on the UI properties of prompt texts.

Input UI Metadata. Through an Android API, we accessed all typing events at full granularity and have access to UI metadata. We log input prompt texts, including titles, labels, and styling/sizing properties of the target text field. Prompt texts are visualized as placeholders and instruct users about what to enter, such as "Message" (WhatsApp) or "Tweet your reply" (X; see Figure 2).

Deriving Contextual Features: Input Intention. We developed a conceptual framework that categorizes prompt texts according to users' *input intention*. By *input intention* we mean the purpose and kind of content that users are supposed to enter into a text field, such as search inputs, direct messaging, or public social media posts. We visualize our concept using the Instagram app as an example in Figure 1. In addition, we present our scheme with seven categories for users' input intention in Table 1.

The Categorization Scheme. We developed a categorization for input text prompts in a multi-stage process. It is based on real mobile typing language data that we collected in a large-scale smartphone sensing study (N = 624, 3 to 6 months). We identified five major intentions that users follow when composing texts on their smartphones: *Messaging*, *Posting*, *Commenting*, *Search*, and *Data Input*. Furthermore, prompt texts that could not be assigned to one of these five categories were labeled as *Other*, or *Ambiguous* if the meaning of a prompt text was unclear. A detailed definition of our categories can be found in Table 1. Our categorization regards all smartphone text entry interactions of all participants, regardless of app category or any other pre-selection. Our finalized schema can be found in the Supplementary materials; we encourage other researchers to use it as a basis to adapt it to the data structure and needs of their own specific research projects. Therefore, we disclose the categorization development process in Section 3.2.

3 DEVELOPMENT AND EVALUATION PROCESS OF OUR INPUT PROMPT TEXT CATEGORIZATION SCHEME

To derive our categories, we applied a semi-automatic, multi-step categorization concept. We therefore used a corpus of real-world input prompt texts that we collected in a smartphone sensing field study. In this section, we describe the collected corpus, the categorization approach, and lastly give some descriptive insights about the resulting categorized data, to give researchers who want to use our procedure some insights about the characteristics of such data.

3.1 Collecting an Input Prompt Text Corpus: Smartphone Sensing Study

We deployed a logging system of smartphone-typed language within a large-scale smartphone sensing study (N = 624, 3 to 6 months) in Germany. This study helped us collect a large body of prompt texts from real-world smartphone interactions to develop a widely applicable categorization scheme for users' input intention. Our corpus contains 10,620 distinct input prompt texts, originating from a total of 1,112,317 text inputs, which obtained a prompt text. We asked our participants to install our self-developed Android app on their smartphones. To access typed language data, this app subscribes to content change events of input fields[4]. Thereby, it is notified of each change, i.e., each typed or removed character. From each character-typing event, we receive a reference to the target input text field[5], containing the metadata of interest, namely the app name that the input field belonged to, and its prompt text. The research app ran continuously in the background of participants' smartphones during the study. Language data were preprocessed with the Language Logger abstraction module of Bemmann and Buschek [3], using the LIWC dictionary[6] [22] for word categorization and a German dictionary [1] as a whitelist for word frequency counting. Thereby, no raw text left the user's device. Data collection was conducted as part of a cooperation project between LMU Munich a the Leibniz Institute for Psychology (ZPID). This research was ethically approved and carefully aligned with EU GDPR guidelines. A detailed description of data collection procedures and collected measures is supplied in the pre-registration of the study protocol [26].

3.2 Prompt Text Categorization Procedure: Distinguishing Language Data By Users' Input Intention

In the first stage, we looked for keywords to categorize text prompts automatically. We decided to classify all prompts that contain German and respective English word snippets "such" / "search" as *Search*, "komment" / "comment" as *Commenting*, and those containing "nachricht" / "message" as *Messaging*. We then categorized all remaining prompt texts manually. We limited our manual categorization to only those prompt texts which occured in more than 0.01% (i.e. 465 prompt texts) of all cases. The categorization was done independently by three researchers (R1 to R3), in three iterations. In the first iteration, R2 and R3 both coded 50 randomly selected prompt texts independently, compared their coding afterwards, discussed differences, and refined the coding scheme. Afterward, R2 and R3 used the refined intention definition to code the remaining prompt texts. Of the 438 manually coded text prompts, they matched in 89.7% of cases (393), and differed in 45 cases. To check the inter-coder reliability, we evaluated Cohen's Kappa [21], resulting in a nearly perfect agreement, according to the classification of Landis and Koch [16] (K = 0.83, 95% CI: $[0.78; 0.88]$). The remaining 45 discrepancies were resolved by R1 who then additionally coded the text prompts. Our categorization labels 983,281 text

[4]Android Accessibility Event type VIEW_TEXT_CHANGED https://developer.android.com/reference/android/view/accessibility/AccessibilityEvent#TYPE_VIEW_TEXT_CHANGED
[5]EditText object https://developer.android.com/reference/android/widget/EditText
[6]LIWC is proprietary software. Its usage must be cleared with the authors.

Table 1: We categorize the *input prompt text* of text inputs on smartphones into *input intentions*

Intention	Description	Example input prompt texts
Messaging	A private message targeted to a defined person or group of people	"Type a message" (WhatsApp), "Enter your message here" (Facebook Messenger)
Posting	(Semi-)public posts in social media applications. They are visible to either anyone (public posts) or limited to a group of people (e.g., friends)	"Write a caption" (Instagram), "What are you doing?" (Facebook)
Commenting	Content that is attached to an existing post, usually with the same visibility as the post	"Comment ..." (Facebook), "Tweet your reply" (X)
Search	Content that constitutes a search query. E.g., inputs into search fields	"Search apps, web, and more..." (Google Quicksearch), "Search photos..." (Gallery app)
Data Input	Inputs that ask the user for some information, usually form fields	"email address" (on a login screen), "Stop, address, ..." (in a public transport service app), "Spanish translation" (in a language learning app)
Other	The input prompt text cannot be assigned exactly one intention, or the purpose is clear but does not belong to one of the five main intentions	E.g., experience sampling and questionnaire items, "write a note..."
Ambiguous	The input prompt text's meaning and purpose are not understandable at all	"0", "???"

inputs from our corpus, which is 88.4% of all text inputs containing an input prompt text. Among all text inputs, including those that do not obtain an input prompt text, our categorization covers 52.63%.

3.3 Descriptive Evaluation: The Characteristics of Input Prompt Text Categorized Data

The following frequencies explain the share of each input intention category among the 983,281 text inputs that were contextualized with our method in the previously described smartphone sensing study.

Input Intention Frequencies. The most frequent input intentions were *Messaging* (44.0%), *Search* (33.8%), and *Data Input* (12.2%). The social media categories *Posting* (1.0%) and *Commenting* (3.2%) occurred less often. The occurrence of the remaining categories *Ambiguous* (4.9%) and *Other* (0.9%) is on a low level, illustrating that our categorization scheme covers the user's input intentions well with nearly 95% of all text inputs.

Text Input Lengths. We found that text inputs that fulfill a functional purpose, such as *Search* ($M = 2.30 words$, $SD = 6.80$) and *Data Input* ($M = 2.73$, $SD = 9.29$), were significantly shorter than texts of intention categories whose content was intended to communicate with other people, such as *Messaging* ($M = 12.43$, $SD = 18.80$), *Posting* ($M = 12.84$, $SD = 19.00$), and *Commenting* ($M = 12.65$, $SD = 20.28$). *Other* text inputs range between both clusters ($M = 5.32$, $SD = 9.42$) (Kruskal-Wallis rank sum test with a consecutive Dunn's test, p-values adjusted with the Bonferroni method. $\alpha < .01$ for all pairwise comparisons).

Data Clarity for Psychological Analyses. We also analyzed the matching rates of resulting smartphone-typed language data in the psychological LIWC dictionary. We found that language data selected with our method, i.e., by input intention, shows higher matching rates and lower standard deviations. We compared the input intention category *Messaging* ($M = 50.64\%$, $SD = 30.33\%$) against data selected by the app category *Communication* ($M = 48.32\%$, $SD = 31.05\%$); and the input intention categories *Posting* ($M = 38.82\%$, $SD = 30.34\%$) and *Comment* ($M = 41.78$, $SD = 30.34\%$) against the app category *Social Media* ($M = 31.59$, $SD = 33.52\%$). Although more in-depth statistical analyses are needed and measures for data quality need to be defined to draw solid conclusions, these preliminary descriptive insights indicate a likely higher data clarity and thus quality of our method compared to previous, traditional methods.

4 LIMITATIONS AND OUTLOOK

First results show that our method of contextualizing smartphone-typed language by user input intention can be valuable in interdisciplinary research. Descriptive statistics indicate that it offers higher data clarity and thus value for research than previous categorization by input target app. We motivate future work to validate our method: Language intentions could be regarded as clusters and evaluated by clustering evaluation metrics such as inter-cluster and intra-cluster distances (c.f. Ranby [24]), or more language-specific similarity metrics as evaluated by Krakovsky and Mokris [15]. Also, it would be interesting to rerun analyses of existing research on mobile language, with both app-categorized data and input prompt text categorized data, to see the effect on the results, for example, changes in effect sizes. Besides supporting research on mobile language use, our method also facilitates intelligent interface concepts in HCI. Our analysis approach of prompt texts and other metadata contributes to an increased understanding of the user's context. Higher context-awareness helps systems to adjust to the users' needs in a particular situation. For example, by understanding an input field's purpose, keyboard word suggestions could be adjusted to better meet expected content and make more specific suggestions. Conversational interaction concepts, especially in the accessibility context, could reach a more fluent user experience if they better understand what the interface expects from its user. Furthermore, large language models could (with sufficient contextual understanding) be used to pre-fill inputs. Thinking beyond our categorization of input motives, one could also think of other kinds of categorizations depending on the specific research goal. Also, further metadata, such as surrounding UI elements, could be taken into account.

Limitations that need to be worked on are the limited availability of input prompt texts. Our implementation could access input prompt texts for 59.5% of all typing sessions. Input prompt texts are not mandatory properties of text fields; developers do not set them for all text inputs. Also privacy aspects of our method need to be carefully considered. Although the more precise filtering of contents that our method enables allows for stricter data minimalism, the privacy implications that language analysis brings need to be tackled. Evaluating users' typed language, no matter how it is preprocessed or abstracted, poses a privacy concern to users and reduces trust in applications and research studies [4].

OPEN SCIENCE

We encourage readers to reproduce and extend our results. Therefore, we made the data used in our study, our analysis scripts, and Android logging libraries available on the Open Science Framework https://osf.io/b62k8.

ACKNOWLEDGMENTS

We would like to thank the Leibniz Institute for Psychology (ZPID) for their great cooperation and support in the Smartphone Sensing Panel Study, from which the dataset published here arises.

This work has been partly supported by the Research Center Trustworthy Data Science and Security (https://rc-trust.ai), one of the Research Alliance centers within the UA Ruhr (https://uaruhr.de).

REFERENCES

[1] 2013. Korpusbasierte Wortgrundformenliste DEREWO, v-ww-bll-320000g-2012-12-31-1.0, mit Benutzerdokumentation. http://www.ids-mannheim.de/derewo

[2] Benjamin Andow, Akhil Acharya, Dengfeng Li, William Enck, Kapil Singh, and Tao Xie. 2017. Uiref: analysis of sensitive user inputs in android applications. In *Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks*. 23–34. https://doi.org/10.1145/3098243.3098247

[3] Florian Bemmann and Daniel Buschek. 2020. LanguageLogger: A Mobile Keyboard Application for Studying Language Use in Everyday Text Communication in the Wild. *Proc. ACM Hum.-Comput. Interact.* 4, EICS, Article 84 (June 2020), 24 pages. https://doi.org/10.1145/3397872

[4] Florian Bemmann and Sven Mayer. 2024. The Impact of Data Privacy on Users' Smartphone App Adoption Decisions. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 278 (Sept. 2024), 23 pages. https://doi.org/10.1145/3676525

[5] Ryan L. Boyd and H. Andrew Schwartz. 2021. Natural Language Analysis and the Psychology of Verbal Behavior: The Past, Present, and Future States of the Field. *Journal of Language and Social Psychology* 40, 1 (2021), 21–41. https://doi.org/10.1177/0261927X20967028

[6] Miriam Brinberg, Nilam Ram, Xiao Yang, Mu-Jung Cho, S Shyam Sundar, Thomas N Robinson, and Byron Reeves. 2021. The idiosyncrasies of everyday digital lives: Using the Human Screenome Project to study user behavior on smartphones. *Computers in Human Behavior* 114 (2021). https://doi.org/10.1016/j.chb.2020.106570

[7] Johannes C Eichstaedt, Margaret L Kern, David B Yaden, H Andrew Schwartz, Salvatore Giorgi, Gregory Park, Courtney A Hagan, Victoria A Tobolsky, Laura K Smith, Anneke Buffone, et al. 2021. Closed-and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods* 26, 4 (2021), 398.

[8] Johannes C. Eichstaedt, Robert J. Smith, Raina M. Merchant, Lyle H. Ungar, Patrick Crutchley, Daniel Preoţiuc-Pietro, David A. Asch, and H. Andrew Schwartz. 2018. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences* 115, 44 (2018), 11203–11208. https://doi.org/10.1073/pnas.1802331115

[9] Sandra González-Bailón, Ning Wang, Alejandro Rivero, Javier Borge-Holthoefer, and Yamir Moreno. 2014. Assessing the bias in samples of large online networks. *Social Networks* 38 (2014), 16–27. https://doi.org/10.1016/j.socnet.2014.01.004

[10] Jianjun Huang, Zhichun Li, Xusheng Xiao, Zhenyu Wu, Kangjie Lu, Xiangyu Zhang, and Guofei Jiang. 2015. {SUPOR}: Precise and scalable sensitive user input detection for android apps. In *24th USENIX Security Symposium (USENIX Security 15)*. 977–992. https://doi.org/10.5555/2831143.2831205

[11] Pascal Jürgens, Birgit Stark, and Melanie Magin. 2020. Two half-truths make a whole? On bias in self-reports and tracking data. *Social Science Computer Review* 38, 5 (2020), 600–615. https://doi.org/10.1177/0894439319831643

[12] Anatoli Kalysch, Davide Bove, and Tilo Müller. 2018. How Android's UI Security is Undermined by Accessibility. In *Proceedings of the 2nd Reversing and Offensive-oriented Trends Symposium*. 1–10. https://doi.org/10.1145/3289595.3289597

[13] Timo K. Koch, Peter Romero, and Clemens Stachl. 2021. Age and gender in language, emoji, and emoticon usage in instant messages. *Computers in Human Behavior* (2021), 106990. https://doi.org/10.1016/j.chb.2021.106990

[14] Timo K. Koch, Peter Romero, and Clemens Stachl. 2022. Age and gender in language, emoji, and emoticon usage in instant messages. *Computers in Human Behavior* 126 (2022), 106990. https://doi.org/10.1016/j.chb.2021.106990

[15] R Krakovsky and I Mokris. 2013. Clustering of text collections based on PART neural network and similarity measure. In *2013 International Conference on System Science and Engineering (ICSSE)*. IEEE, 253–257. https://doi.org/10.1109/ICSSE.2013.6614669

[16] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174. https://doi.org/10.2307/2529310

[17] Hansoo Lee, Joonyoung Park, and Uichin Lee. 2021. A Systematic Survey on Android API Usage for Data-Driven Analytics with Smartphones. *arXiv preprint arXiv:2104.11271* (2021). https://doi.org/10.1145/3530814

[18] Xuan Lu, Wei Ai, Xuanzhe Liu, Qian Li, Ning Wang, Gang Huang, and Qiaozhu Mei. 2016. Learning from the Ubiquitous Language: An Empirical Analysis of Emoji Usage of Smartphone Users *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 770–780. https://doi.org/10.1145/2971648.2971724

[19] Elizabeth A McNeilly, Kathryn L Mills, Lauren E Kahn, Ryann Crowley, Jennifer H Pfeifer, and Nicholas B Allen. 2023. Adolescent social communication through smartphones: Linguistic features of internalizing symptoms and daily mood. *Clinical Psychological Science* 11, 6 (2023), 1090–1107. https://doi.org/10.1177/21677026221125180

[20] Joshua Ofoeda, Richard Boateng, and John Effah. 2019. Application programming interface (API) research: A review of the past to inform the future. *International Journal of Enterprise Information Systems (IJEIS)* 15, 3 (2019), 76–95. https://doi.org/10.4018/ijeis.2019070105

[21] Cliodhna O'Connor and Helene Joffe. 2020. Intercoder reliability in qualitative research: debates and practical guidelines. *International journal of qualitative methods* 19 (2020), 1609406919899220. https://doi.org/10.1177/1609406919899220

[22] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. The Development and Psychometric Properties of LIWC2015. (2015).

[23] Nataniel Pereira Borges Junior. 2020. Learning the language of apps. (2020). https://doi.org/10.22028/D291-32884

[24] Erik Ranby. 2016. A comparison of clustering techniques for short social text messages.

[25] André Rodrigues, André R.B. Santos, Kyle Montague, Hugo Nicolau, and Tiago Guerreiro. 2021. WildKey: A Privacy-Aware Keyboard Toolkit for Data Collection In-The-Wild. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers* (Virtual, USA) *(UbiComp/ISWC '21 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 542–545. https://doi.org/10.1145/3460418.3482872

[26] Ramona Schoedel and Michelle Oldemeier. 2020. Basic Protocol: Smartphone Sensing Panel Study. (2020). https://doi.org/10.23668/PSYCHARCHIVES.2901

[27] Ramona Schoedel, Michelle Oldemeier, Léonie Bonauer, and Larissa Sust. 2022. Dataset for: Systematic categorisation of 3091 smartphone applications from a large-scale smartphone sensing dataset. https://doi.org/10.23668/psycharchives.5362

[28] H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* 8, 9 (2013), e73791. https://doi.org/10.1371/journal.pone.0073791

[29] Clemens Stachl, Quay Au, Ramona Schoedel, Samuel D Gosling, Gabriella M Harari, Daniel Buschek, Sarah Theres Völkel, Tobias Schuwerk, Michelle Oldemeier, Theresa Ullmann, et al. 2020. Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences* 117, 30 (2020), 17680–17687. https://doi.org/10.1073/pnas.1920484117

[30] Lieke Verheijen and Wessel Stoop. 2016. Collecting Facebook Posts and WhatsApp Chats: Corpus Compilation of Private Social Media Messages. In *Text, Speech, and Dialogue*. Springer International Publishing, Cham, 249–258. https://doi.

org/10.1007/978-3-319-45510-5_29

[31] Rafael Wampfler, Severin Klingler, Barbara Solenthaler, Victor R. Schinazi, Markus Gross, and Christian Holz. 2022. Affective State Prediction from Smartphone Touch and Sensor Data in the Wild. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 403, 14 pages. https://doi.org/10.1145/3491102.3501835

[32] Tanapuch Wanwarang, Nataniel P. Borges, Leon Bettscheider, and Andreas Zeller. 2020. Testing Apps With Real-World Inputs. In *Proceedings of the IEEE/ACM 1st International Conference on Automation of Software Test* (Seoul, Republic of Korea) *(AST '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3387903.3389310

Table 2: In this table, we disclose details on our categorization process, and show statistics on inter-rater agreement according to O'Connor and Joffe [21]. Number of assigned intention categories, alongside disagreements and inter-rater agreement to each intention category of the manual coding process.

	Assigned Prompt Texts				
Input Intention	**Overall**	**Manually Coded**	**Disagreements**	**Cohen's Kappa**	**95% CI**
Messaging	568	31	4	0.88	[0.79;0.97]
Posting	20	20	4	0.75	[0.6;0.9]
Commenting	157	4	1	0.75	[0.41;1]
Search	2997	16	1	0.78	[0.63;0.93]
Data Input	248	248	12	0.85	[0.81;0.9]
Other	13	13	7	0.8	[0.57;1]
Ambiguous	105	105	16	0.82	[0.75;0.88]

Table 3: Comparing characteristics of text inputs filtered by input intention (yellow background) and app category (green background). We compare mean and standard deviation for the two variables *matching rate* and *number of words per text input*. A Kruskal-Wallis rank sum test with a consecutive Dunn's test (p-values adjusted with the Bonferroni method) revealed significant differences between the three groups *Search* and *Data Input* vs. *Messaging, Posting* and *Commenting* vs. *Other* for all pairwise comparisons with $\alpha < .01$.

	LIWC matching rate		words per text input	
Input Text Selection	**M**	**SD**	**M**	**SD**
App Category: Communication	48.32%	31.05%	16.43	38.30
Input Intention: Messaging	50.64%	30.33%	12.43	18.80
App Category: Social Media	31.59%	33.52%	12.87	39.60
Input Intention: Posting	38.82%	30.34%	12.84	19.00
Input Intention: Commenting	41.78%	30.35%	12.65	20.28