What Do Dialect Speakers Want?
A Survey of Attitudes Towards Language Technology for German Dialects

Verena Blaschke⬠ **Christoph Purschke**⬟ **Hinrich Schütze**▲⬠ **Barbara Plank**▲⬠⊘

▲ Center for Information and Language Processing (CIS), LMU Munich, Germany
⬠ Munich Center for Machine Learning (MCML), Munich, Germany
⬟ Department of Humanities, University of Luxembourg, Luxembourg
⊘ Department of Computer Science, IT University of Copenhagen, Denmark
{verena.blaschke, b.plank}@lmu.de

Abstract

Natural language processing (NLP) has largely focused on modelling standardized languages. More recently, attention has increasingly shifted to local, non-standardized languages and dialects. However, the relevant speaker populations' needs and wishes with respect to NLP tools are largely unknown. In this paper, we focus on dialects and regional languages related to German – a group of varieties that is heterogeneous in terms of prestige and standardization. We survey speakers of these varieties (N=327) and present their opinions on hypothetical language technologies for their dialects. Although attitudes vary among subgroups of our respondents, we find that respondents are especially in favour of potential NLP tools that work with dialectal input (especially audio input) such as virtual assistants, and less so for applications that produce dialectal output such as machine translation or spellcheckers.

1 Introduction

Most natural language processing (NLP) research focuses on languages with many speakers, high degrees of standardization and large amounts of available data (Joshi et al., 2020). Only recently, more NLP projects have started to include local, non-standardized languages and dialects. However, different speakers and cultures have different needs. As recently echoed by multiple researchers, the creation of language technologies (LTs) should take into account what the relevant speaker community finds useful (Bird, 2020, 2022; Liu et al., 2022; Mukhija et al., 2021), and communities can differ from one another in that regard (Lent et al., 2022).

In this work, we focus on dialects and regional languages[1] closely related to German (for the sake of simplicity, we use 'dialects' to refer to these varieties in this paper). With dialect competence generally being in decline in the German-speaking area,

---

[1]Our survey also includes responses by speakers of Low German, which is officially recognized as a regional language.
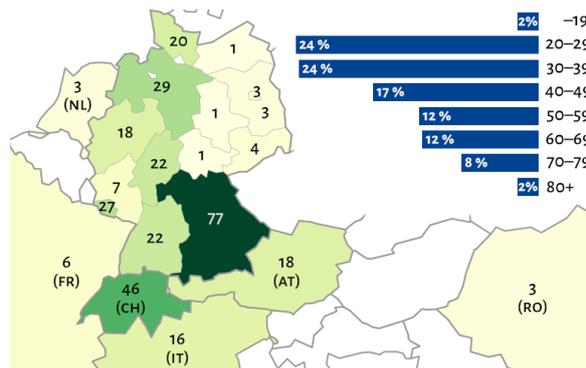

Figure 1: **Countries and German states in which the respondents' dialects are spoken**, with the number of respective respondents, and the overall age distribution.

today, dialect speakers usually also speak Standard German, while dialects often are replaced by regiolects – intermediate varieties between standard and dialect (Kehrein, 2019). Speaker attitudes towards dialects vary greatly (Gärtig et al., 2010, pp. 155–167).

Although these dialects are predominantly spoken and only few of them have traditional orthographies, many of them are also used in written, digital contexts (Androutsopoulos, 2003). Accordingly, some NLP datasets based (primarily) on such digital data exist, and a small number is also annotated for NLP tasks (Blaschke et al., 2023). Several recent publications feature LTs for German dialects, such as machine translation (Haddow et al., 2013; Honnet et al., 2018; Lambrecht et al., 2022; Aepli et al., 2023a; Her and Kruschwitz, 2024), speech-to-text (Herms et al., 2016; Nigmatulina et al., 2020; Gerlach et al., 2022) and text-to-speech systems (Gutscher et al., 2023), and slot and intent detection for conversational assistants (van der Goot et al., 2021; Aepli et al., 2023b; Winkler et al., 2024; Abboud and Oz, 2024).

To investigate what speaker communities are interested in, we survey dialect speakers from different German-speaking areas (Figure 1). We gather

a snapshot of their current attitudes towards LTs to answer the following questions: *Q1)* Which dialect technologies do respondents find especially useful (§4.2)? *Q2)* Does this depend on whether the in- or output is dialectal, and on whether the LT works with speech or text data (§4.3)? *Q3)* How does this reflect relevant sociolinguistic factors (§4.4)?

2 Related Work

The closest survey to ours on investigating attitudes of speakers of non-standard language varieties towards LTs is by Lent et al. (2022). They conducted a survey on the actual and desired LT use by speakers of different creoles (*N*=37). They find that the needs vary from speaker community to speaker community, and that speakers who are also highly proficient in the local high-prestige language are less interested in creole LTs. Of the technologies included in the survey, speech-related technologies (transcription and synthesis) are the most popular; machine translation (MT) and question answering software are also desired by multiple communities, while spellcheckers are controversial.

Soria et al. (2018) surveyed speakers of four regional European languages[2] about whether and why they use (or do not use) their languages in digital contexts. When asked about the desirability of currently unavailable spellcheckers and MT systems, more respondents judged both as desirable than not, although the exact proportions vary across communities. Millour (2019; 2020, pp. 230, 239) found similar results in surveys of Mauritian Creole (*N*=144) and Alsatian speakers (*N*=1,224).

Conversely, Way et al. (2022) investigate actual LT use by speakers of different European national languages (91–922 respondents per country). The most commonly used LTs are MT, search engines and spell- or grammar checkers. When respondents do not use specific LTs, this can simply be due to the absence of such technologies for certain languages, but also due to a lack of interest in specific language–LT combinations.

Recently, several surveys have also investigated speaker community perspectives regarding LTs for many different indigenous language communities (Mager et al., 2023; Cooper et al., 2024; Dolinska et al., 2024; Tonja et al., 2024). However, these surveys focus on languages with very different sociolinguistic contexts than the ones in our survey and

that are unrelated to their respective local high-resource languages.

3 Questionnaire

Our questionnaire is aimed at speakers of German dialects and related regional languages and consists of two main parts: We ask our participants about their dialect, and we ask about their opinions on hypothetical LTs for their dialect. Several of the questions regarding dialect use are inspired by Soria et al. (2018) and Millour (2020), and we choose a similar selection of LTs as Way et al. (2022) (§4.2). For each technology, we provide a brief definition to make the survey accessible to a broad audience (e.g., 'Speech-to-text systems transcribe spoken language. They are for instance used for automatically generating subtitles or in the context of dictation software.'). We then ask participants to rate on a 5-point Likert scale how useful they would find such a tool for their dialect. We allow respondents to elaborate on their answers in comment fields. The full questionnaire is in Appendix §A.

The questionnaire was written in German, and was estimated to take between 10–15 minutes for completion.[3] It was online for three weeks in September and October 2023 and got disseminated via word of mouth, social media, mailing lists and by contacting dialect and heritage societies. Our results are hence based on a convenience sample.

4 Results

We reached 441 people, 327 of whom are self-reported dialect speakers and finished the entire questionnaire – their responses are presented in the following. Detailed answer distributions are in Appendix §A; correlations are in §B.

4.1 Dialect Background and Attitudes

Most of our respondents answer that they have a very good command of their dialect (68%) and acquired it as a mother tongue (71%). Figure 1 shows where the respondents' dialects are spoken (and their age distribution): mostly in Germany (72%), followed by Switzerland (14%) and Austria (6%).[4] Nearly a quarter (24%) each are in

---

[2]Karelian (*N*=156, Salonen, 2017), Breton (*N*=202, Hicks, 2017), Basque (*N*=427, Gurrutxaga Hernaiz and Ceberio Berger, 2017), Sardinian (*N*=516, Russo and Soria, 2017).

[3]80% of our respondents took <15 min to fill out the entire questionnaire, and 60% even less than 10 min.

[4]The other varieties are spoken in areas with minority speaker communities: Italy (Bavarian), France (Alemannic), the Netherlands (Low German), and Romania. The geographic distribution of our respondents is not representative of the overall dialect speaker population.

their twenties and thirties, almost all others are older. When rating how traditional their dialect is on a scale from 1 (traditional dialect that people from other regions have trouble understanding) to 5 (regionally marked German easily understood by outsiders), the largest group of respondents (35%) indicated a 2 (μ=2.6, σ=1.1).

Just over half of our respondents (52%) speak their dialect on a daily basis, and 43% indicate that they would like to use their dialect in all areas of life. Most respondents (70%) value the diversity of their dialect. Nearly two thirds (65%) are opposed to having a standardized orthography for their dialect. Just over half of the respondents (53%) say that their dialect is only spoken and not well-suited for written communication – nevertheless, two thirds (66%) also write their dialect, even if rarely. Many (63%) find it easy to read dialectal texts written by others. Written dialect is commonly used for communicating with others – the most common writing scenarios are text messages (57%, multiple responses possible), followed by letters/emails (26%), social media posts and comments (19% each) – but also for notes to oneself including diary entries (19%).

About a third (35%) indicate that they are actively engaged in dialect preservation pursuits (multiple responses possible): 13% as members of dialect preservation societies, 4% as teachers, and 22% in other ways. Write-in comments by the last group point out other language-related professions, but also include speaking the dialect in public or with children as a means of active dialect preservation.[5] We compare the opinions of respondents with and without such dialect engagement in §4.4.

14% of our respondents are familiar with at least one LT that already caters to their dialect. Just over half of the respondents (54%) indicate that their dialect being represented by more LTs would make it more attractive to younger generations, and a smaller group (31%) says they would use their dialect more often given suitable LTs.

---

[5]The write-in answers contain 13 mentions of speaking the dialect with family members (especially children and grandchildren), 18 mentions of simply speaking the dialect (in public), 14 mentions of carrying out dialect-related research (as a job or a hobby), and 10 mentions of using the dialect in the context of literature or music, with slight overlap between these groups. None of these subgroups are concentrated in any specific area, but instead include respondents from areas where dialects and regional languages have very different statuses (cf. §4.4): Low German speakers as well as other German respondents, Swiss respondents, respondents from countries where German is a minority language, and so on.
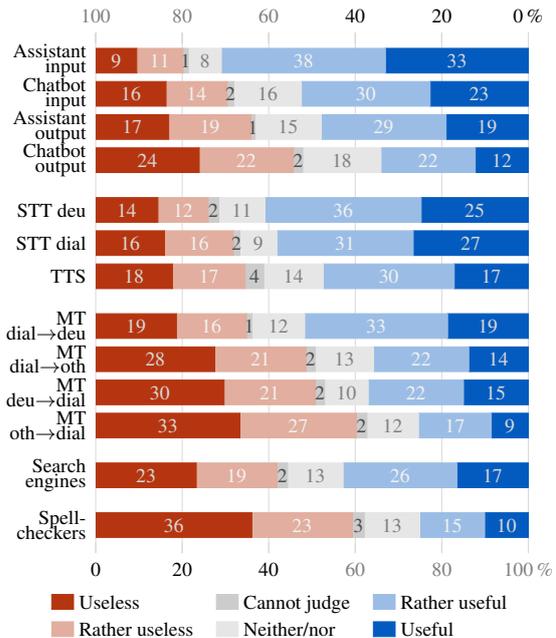

Figure 2: **Opinions on potential language technologies for dialects.** *STT*=speech-to-text, *TTS*=text-to-speech, *dial*=dialect, *deu*=German, *oth*=other languages, *MT*=machine translation, *cannot judge*=skip question.

4.2 Which dialect LTs are deemed useful?

Figure 2 shows our respondents' opinions on LTs (*Q1*), and Appendix §C presents the average scores per LT when responses are mapped to a numerical scale. While there are diverging opinions on every LT – there is no single technology that (nearly) all respondents consider useful or useless for their dialect – some trends emerge, as we discuss next.[6] Overall, the responses are generally correlated with each other: respondents who think positively/negatively of one technology tend to think similarly about others. Nevertheless, some LTs are overall more popular, and some less so:

Virtual assistants and chatbots The most clearly favoured LT by dialect speakers in our survey are virtual assistants (such as Siri or Alexa) that can respond to dialectal input (71% in favour, 20% against). Chatbots that can handle dialectal input are less popular, but still deemed useful by a slight majority (52%). Systems that could output dialectal responses are less popular: 48% would

---

[6]Of the technologies we included, MT, search engines and spell checkers are the most used LTs in the EU (Way et al., 2022, p. 26). We assume that the tools that people use a lot are also tools they generally deem useful, yet those tools are all ranked relatively low in our results – suggesting that our results reveal attitudes on dialect LTs rather than LTs in general.

find virtual assistants that answer in dialect useful, and 34 % think so about chatbots.

Speech-to-text and text-to-speech When asked about speech-to-text (STT) software, a majority (61 %) is in favour of systems that transcribe spoken dialect into written Standard German, and a slightly smaller majority (58 %) is in favour of written dialectal output. When it comes to text-to-speech (TTS) systems that synthesize dialect text into a spoken form, the respondents are even more split, with 47 % in favour and 35 % against.

Machine translation We ask for opinions on four different configurations regarding automatic translation of written texts: each possible combination for translation into vs. out of the dialect and from/into Standard German vs. a foreign language. All options are to some degree controversial among the respondents, with translation from the dialect into Standard German being the most popular (52 % in favour) and from the dialect into a foreign language the least popular (25 % in favour).

Search engines Search engines that could deal with dialectal input are controversial, with 43 % each in favour of and against this LT, although the negative group holds stronger opinions. Some write-in comments question whether (monolingual) information retrieval would produce useful results or mention finding it easier to write in Standard German rather than in a dialect, but others voice a desire to be able to find results for queries including dialectal terms with no direct German equivalent.

Spellcheckers Most respondents (59 %) are opposed to spell- or grammar checkers for their dialect, although a quarter (25 %) is in favour. Several respondents mention opposition to spellcheckers since they want their dialectal writing to exactly reflect the pronunciation and word choices of their local dialect and would be bothered if a spellchecker changed them to an arbitrary standardized version of the dialect.

4.3 Are there differences for dialect input vs. output and text vs. speech?

As seen in the previous section, there is a general tendency to prefer versions of LTs that process dialectal input rather than produce dialectal output (*Q2*). Several write-in comments voice worries about dialectal output not modelling their dialect accurately enough. Additionally, technologies dealing with spoken language tend to be rated more positively than those focusing on text only.

Correlation with attitudes towards orthography Being in favour of a standardized dialect orthography is positively, albeit not very strongly, correlated with being in favour of any technology involving a written version of the dialect and/or (written or spoken) dialectal output (Spearman's ρ values between 0.14 and 0.47 per LT with p-values <0.001).

4.4 Do attitudes reflect sociolinguistic factors?

To address *Q3* and the heterogeneity of our respondent group, we compare answers between larger subgroups. We summarize the results of t-tests with p-values <0.05. Appendix §D provides more details, together with two additional comparisons that only have small effect sizes (speaker age and dialect traditionality).

Language activists Since language activists might have overly enthusiastic attitudes compared to the speaker population at large (Soria et al., 2018), we compare those who report involvement in dialect preservation ('activists', *N*=115) to those who do not (*N*=212). Activists are generally more in favour of LTs for dialects, with statistically significant differences for (any kind of) machine translation, TTS software, spellcheckers, and search engines, as well as for written dialect output options for STT, chatbots and virtual assistants. Removing the activists' responses from our analysis only barely changes the order of preferred LTs (§C).

Region Additionally, we compare three large regional subgroups with different sociolinguistic realities. In Germany and Austria, traditional dialects have been partially replaced by more standard-like regiolects, while dialects have high prestige in Switzerland where Standard German is often reserved for writing (Kehrein, 2019; Ender and Kaiser, 2009). Low German, traditionally spoken in parts of Northern Germany and the Eastern Netherlands, is officially recognized as a language and is more distantly related to Standard German than the other varieties our participants speak. Its speaker numbers are in decline, but many Northern Germans think Low German should receive more support in, e.g., public schools (Adler et al., 2016).

We compare the opinions of Swiss (*N*=46) and Low German (*N*=58) respondents to German and Austrian non-Low-German speakers (*N*=191).[7]

---

[7]We identify the Low German respondents based on the

Our Low German respondents are more in favour of a standardized orthography and of spellcheckers than our other German and Austrian respondents, the Swiss respondents less so. This is unsurprising in that several orthographies have been proposed for Low German, whereas (typically spoken) dialects and (typically written) Standard German exist in a diglossic state in Switzerland. Nevertheless, *both* groups are more in favour of STT software with dialectal output. The Low German respondents are more in favour of chatbots with dialectal answers, TTS, (any kind of) MT and search engines. Swiss Germans rate virtual assistants that can handle dialectal input as more desirable (87 % in favour), and are more in favour of STT software with Standard German output.

5 Discussion and Conclusion

We surveyed speakers of dialect varieties on their attitudes towards LTs. Generally, the survey participants prefer LTs working with dialect input rather than output. They also tend to prefer tools that process speech over those for text (*Q2*). This is consistent with Chrupała's (2023) argument that NLP should focus more on spoken language to better represent actual language use. It also reflects the complex, often conflicting attitudes speakers of multiple varieties have towards competing linguistic and social norms. Consequently, the most popular potential dialect LTs (*Q1*) in our survey process spoken dialectal input: virtual assistants with dialect input and speech-to-text systems.

However, like Lent et al. (2022), we find that different speaker communities vary in their attitudes towards LTs (*Q3*). For instance, opinions on the standardization of a dialect are a relevant factor regarding the desirability of written LTs. Nevertheless, the acceptance and rejection of LTs is related to individual factors beyond just attitudes, e.g., experience with and trust in digital technology.

We hope that our study inspires other NLP researchers to actively consider the wants and needs of the relevant speaker communities. Based on the results of this study, we also encourage the dialect NLP community to pursue more work on spoken language processing.

---

dialect they indicated speaking (§A2), combined with region information for respondents who supplied ambiguous dialect names. For the (other) German, Austrian, and Swiss respondents, we used region information.

Ethical Considerations

We only collected responses from participants who consented to having their data saved and analyzed for research purposes. We did not ask for personally identifying information. We store responses on a local server and only share results based on aggregate analyses. Appendix §A contains the full questionnaire including the introduction where we describe the purpose of the study and explain what data we collect and how we use the data.

Participation was completely anonymous, voluntary and with no external reward. We do not see any particular risks associated with this work.

Limitations

Our results are based on a convenience sample; neither the geographic or age distribution are representative of the population at large (dialect-speaking or not). Language activists are over-represented (hence our additional analysis in §4.4 and Appendices §C and §D), and participating in the survey may have been especially of interest to people who feel (in one way or another) strongly about the topic of dialects and technology. Even so, our respondents are not a monolith in their opinions and we can see meaningful differences between the relative popularity of different technologies.

We aimed to keep participation effort low and therefore limited the number of questions we included. We considered asking "Would you use *X* if it existed?" in addition to "Would you find *X* useful?" to explicitly disentangle the participants' own needs from what are possibly the perceived needs of the community. We decided against this in order to keep the questionnaire as short as possible and because we were unsure how accurate such assessment would be.

The scale of our answer possibilities uses "not useful" as the opposite of "useful." However, it would be interesting to instead use a scale from "harmful" to "useful" in future surveys, in order to get a better impression of whether respondents who deem an LT useless in our version of the survey find it actively harmful or merely uninteresting.

To minimize the total time needed to fill out the questionnaire and to guarantee the privacy of the respondents after asking respondents about what specific dialect they speak (used later to identify the Low German speakers), we intentionally kept additional demographic questions at a minimum and did not ask about education, income, gender,

or similar variables.

As this survey is based on self-reporting, we expect discrepancies between reported and actual opinions and behaviour. Since participation was anonymous and entirely voluntary with no external reward, we think it unlikely for participants to lie about their opinions. It is likely, though, that (especially younger) participants overstate their dialect competence or the traditionality of their dialect, in line with overall dialect dynamics in German (Purschke, 2011).

Acknowledgements

References

Khadige Abboud and Gokmen Oz. 2024. Towards equitable natural language understanding systems for dialectal cohorts: Debiasing training data. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 16487–16499, Torino, Italia. ELRA and ICCL.

Astrid Adler, Christiane Ehlers, Reinhard Goltz, Andrea Kleene, and Albrecht Plewnia. 2016. Status und Gebrauch des Niederdeutschen 2016. Erste Ergebnisse einer repräsentativen Erhebung. Technical report, Institut für Deutsche Sprache, Mannheim.

Noëmi Aepli, Chantal Amrhein, Florian Schottmann, and Rico Sennrich. 2023a. A benchmark for evaluating machine translation metrics on dialects without standard orthography. In *Proceedings of the Eighth Conference on Machine Translation*, pages 1045–1065, Singapore. Association for Computational Linguistics.

Noëmi Aepli, Çağrı Çöltekin, Rob Van Der Goot, Tommi Jauhiainen, Mourhaf Kazzaz, Nikola Ljubešić, Kai North, Barbara Plank, Yves Scherrer, and Marcos Zampieri. 2023b. Findings of the VarDial evaluation campaign 2023. In *Tenth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2023)*, pages 251–261, Dubrovnik, Croatia. Association for Computational Linguistics.

Jannis Androutsopoulos. 2003. Online-Gemeinschaften und Sprachvariation. Soziolinguistische Perspektiven auf Sprache im Internet. *Zeitschrift für germanistische Linguistik*, 31(2).

Steven Bird. 2020. Decolonising speech and language technology. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3504–3519, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Steven Bird. 2022. Local languages, third spaces, and other high-resource scenarios. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7817–7829, Dublin, Ireland. Association for Computational Linguistics.

Verena Blaschke, Hinrich Schütze, and Barbara Plank. 2023. A survey of corpora for Germanic low-resource languages and dialects. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 392–414, Tórshavn, Faroe Islands. University of Tartu Library.

Grzegorz Chrupała. 2023. Putting natural in Natural Language Processing. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7820–7827, Toronto, Canada. Association for Computational Linguistics.

Ned Cooper, Courtney Heldreth, and Ben Hutchinson. 2024. "it's how you do things that matters": Attending to process to better serve indigenous communities with language technologies. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 204–211, St. Julian's, Malta. Association for Computational Linguistics.

Joanna Dolinska, Shekhar Nayak, and Sumittra Suratdecha. 2024. Akha, dara-ang, karen, khamu, Mlabri and urak lawoi' language minorities' subjective perception of their languages and the outlook for development of digital tools. In *Proceedings of the Seventh Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 94–99, St. Julians, Malta. Association for Computational Linguistics.

Andreas Ender and Irmgard Kaiser. 2009. Zum Stellenwert von Dialekt und Standard im österreichischen und Schweizer Alltag: Ergebnisse einer Umfrage. *Zeitschrift für germanistische Linguistik*, 37(2):266–295.

Anne-Kathrin Gärtig, Albrecht Plewnia, and Astrid Rothe. 2010. *Wie Menschen in Deutschland über Sprache denken. Ergebnisse einer bundesweiten Repräsentativerhebung zu aktuellen Spracheinstellungen*. Institut für Deutsche Sprache, Mannheim.

Johanna Gerlach, Jonathan Mutal, and Bouillon Pierrette. 2022. Producing Standard German subtitles for Swiss German TV content. In *Ninth Workshop on Speech and Language Processing for Assistive Technologies (SLPAT-2022)*, pages 37–43, Dublin, Ireland. Association for Computational Linguistics.

Antton Gurrutxaga Hernaiz and Klara Ceberio Berger. 2017. Basque – a digital language? Technical report. Reports on Digital Language Diversity in Europe.

Lorentz Gutscher, Michael Pucher, and Víctor Garcia. 2023. Neural speech synthesis for Austrian dialects with standard German grapheme-to-phoneme conversion and dialect embeddings. In *Proceedings of the 2nd Annual Meeting of the Special Interest Group on Under-resourced Languages (SIGUL 2023)*.

Barry Haddow, Adolfo Hernández, Friedrich Neubarth, and Harald Trost. 2013. Corpus development for machine translation between standard and dialectal varieties. In *Proceedings of the Workshop on Adaptation of Language Resources and Tools for Closely Related Languages and Language Variants*, pages 7–14, Hissar, Bulgaria. INCOMA Ltd. Shoumen, BULGARIA.

Wan-Hua Her and Udo Kruschwitz. 2024. Investigating neural machine translation for low-resource languages: Using Bavarian as a case study. In *Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-resourced Languages*.

Robert Herms, Laura Seelig, Stefanie Münch, and Maximilian Eibl. 2016. A corpus of read and spontaneous Upper Saxon German speech for ASR evaluation. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4648–4651, Portorož, Slovenia. European Language Resources Association (ELRA).

Davyth Hicks. 2017. Breton – a digital language? Technical report. Reports on Digital Language Diversity in Europe.

Pierre-Edouard Honnet, Andrei Popescu-Belis, Claudiu Musat, and Michael Baeriswyl. 2018. Machine translation of low-resource spoken dialects: Strategies for normalizing Swiss German. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

Roland Kehrein. 2019. Vertical language change in Germany: Dialects, regiolects, and standard German. In Stanley D. Brunn and Roland Kehrein, editors, *Handbook of the Changing World Language Map*. Springer International Publishing.

Louisa Lambrecht, Felix Schneider, and Alexander Waibel. 2022. Machine translation from standard German to Alemannic dialects. In *Proceedings of the 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages*, pages 129–136, Marseille, France. European Language Resources Association.

Heather Lent, Kelechi Ogueji, Miryam de Lhoneux, Orevaoghene Ahia, and Anders Søgaard. 2022. What a creole wants, what a creole needs. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6439–6449, Marseille, France. European Language Resources Association.

Zoey Liu, Crystal Richardson, Richard Hatcher, and Emily Prud'hommeaux. 2022. Not always about you: Prioritizing community needs when developing endangered language technology. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3933–3944, Dublin, Ireland. Association for Computational Linguistics.

Manuel Mager, Elisabeth Mager, Katharina Kann, and Ngoc Thang Vu. 2023. Ethical considerations for machine translation of indigenous languages: Giving a voice to the speakers. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4871–4897, Toronto, Canada. Association for Computational Linguistics.

Alice Millour. 2019. Getting to know the speakers: a survey of a non-standardized language digital use. In *9th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznań, Poland.

Alice Millour. 2020. *Myriadisation de ressources linguistiques pour le traitement automatique de langues non standardisées*. Ph.D. thesis, Sorbonne Université.

Namrata Mukhija, Monojit Choudhury, and Kalika Bali. 2021. Designing language technologies for social good: The road not taken. *Computing Research Repository*, 2110.07444.

Iuliia Nigmatulina, Tannon Kew, and Tanja Samardzic. 2020. ASR for non-standardised languages with dialectal variation: the case of Swiss German. In *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects*, pages 15–24, Barcelona, Spain (Online). International Committee on Computational Linguistics (ICCL).

Christoph Purschke. 2011. *Regionalsprache und Hörerurteil. Grundzüge einer perzeptiven Variationslinguistik*. Steiner, Stuttgart.

Irene Russo and Claudia Soria. 2017. Sardinian – a digital language? Technical report. Reports on Digital Language Diversity in Europe.

Tuomo Salonen. 2017. Karelian – a digital language? Technical report. Reports on Digital Language Diversity in Europe.

Jürgen Erich Schmidt, Joachim Herrgen, Roland Kehrein, Alfred Lameli, and Hanna Fischer. 2020–. Regionalsprache.de (REDE III. Forschungsplattform zu den modernen Regionalsprachen des Deutschen. Edited by Robert Engsterhold, Heiko Girnth, Simon Kasper, Juliane Limper, Georg Oberdorfer, Tillmann Pistor, Anna Wolańska. Assisted by Dennis Beitel, Milena Gropp, Maria Luisa Krapp, Vanessa Lang, Salome Lipfert, Jeffrey Pheiff, Bernd Vielsmeier.

Claudia Soria, Valeria Quochi, and Irene Russo. 2018. The DLDP survey on digital use and usability of EU regional and minority languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Atnafu Lambebo Tonja, Fazlourrahman Balouchzahi, Sabur Butt, Olga Kolesnikova, Hector Ceballos, Alexander Gelbukh, and Thamar Solorio. 2024. NLP progress in Indigenous Latin American languages.

Rob van der Goot, Ibrahim Sharaf, Aizhan Imankulova, Ahmet Üstün, Marija Stepanović, Alan Ramponi, Siti Oryza Khairunnisa, Mamoru Komachi, and Barbara Plank. 2021. From masked language modeling to translation: Non-English auxiliary tasks improve zero-shot spoken language understanding. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2479–2497, Online. Association for Computational Linguistics.

Andy Way, Georg Rehm, Jane Dunne, Jan Hajič, Teresa Lynn, Maria Giagkou, Natalia Resende, Tereza Vojtěchová, Stelios Piperidis, Andrejs Vasiljevs, Aivars Berzins, Gerhard Backfried, Marcin Skowron, Jose Manuel Gomez-Perez, Andres Garcia-Silva, Martin Kaltenböck, and Artem Revenko. 2022. Report on all external consultations and surveys. Technical report, European Language Equality.

Miriam Winkler, Virginija Juozapaityte, Rob van der Goot, and Barbara Plank. 2024. Slot and intent detection resources for Bavarian and Lithuanian: Assessing translations vs natural queries to digital assistants. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 14898–14915, Torino, Italy. ELRA and ICCL.

A Questionnaire

In this section, we reproduce the questions and answers from our survey, in the original wording as well as in translation. Translations are in *grey italics,* remarks about the questionnaire are in *black italics.* Answer options that end with a colon (:) came with an optional text input field in the questionnaire. All questions except for the first two could be skipped without answering.

———

Herzlich willkommen, servus, grüezi & moin!

Sprachtechnologie ist momentan allgegenwärtig, ob bei Übersetzungsprogrammen, Chatbots oder anderen Anwendungen. Hauptsächlich unterstützen diese Anwendungen lediglich Standardsprachen – was nicht unbedingt dem entspricht, wie wir im Alltag Sprache verwenden.

Daher möchten wir herausfinden, wie Sie als Sprecher*innen von Dialekten und Regionalsprachen möglicher Sprachtechnologie für Ihre Sprachform gegenüberstehen: **welche Anwendungen halten Sie für wünschenswert bzw. unnötig?**

Welcome and hello [in different dialects]!

Language technology is currently omnipresent – be it in the context of translation software, chatbots or other applications. Such applications primarily support standard languages – which is not necessarily how we use language in our everyday lives.

Because of this we would like to find out what you as speakers of dialects and regional languages think of potential technologies for your language variety: which applications do you find desirable or useless?

———

Das Ausfüllen des Fragebogens dauert etwa 10–15 Minuten.

Wir behandeln Ihre Antworten vertraulich und veröffentlichen diese nur in anonymisierter Form und ohne dass Rückschlüsse auf Ihre Person gezogen werden können.

Genauere Details:

Ziel der Befragung ist es zum einen, herauszufinden, ob es Unterschiede zwischen den Arten von Sprachtechnologien gibt, die Dialektsprecher*innen tendenziell als nützlich bzw. nutzlos bewerten. Zum anderen möchten wir herausfinden, ob ein statistischer Zusammenhang zwischen diesen Antworten und dem Dialekthintergrund und -gebrauch der Befragten besteht.

Die Antworten werden auf einem Server der LMU in München gespeichert. Wir speichern dabei nur Ihre Antworten und den Antwortzeitpunkt (um die typische Ausfülldauer besser einzuschätzen), nicht aber Ihre IP-Adresse. Wir geben die Daten *nicht* an Dritte weiter, sondern veröffentlichen lediglich Ergebnisse, die auf Aggregatdaten und statistischen Analysen beruhen. Zudem zitieren wir gegebenenfalls aus (optional gegebenen) Kommentarfeld-Antworten.

Kontaktmöglichkeit bei Fragen oder Kommentaren zu dieser Umfrage: *[Contact data of first author].*

Vielen Dank für Ihre Teilnahme!

This questionnaire takes about 10–15 minutes to fill out.

We treat your answers as confidential and only share them as anonymized data that do not allow drawing any inferences about your identity.

More detailed information:

The goal of this survey is firstly to determine whether there are differences between the types of language technologies that dialect speakers tend to find useful or useless. Additionally, we would like to find out whether there is a statistical correlation between the answers and the dialect background of the participants.

We store the answers on an LMU server in Munich. This only includes storing your answers and the time of the questions are answered (to better estimate the typical response duration), but not your IP address. We do not *share your data with third parties, but only share results based on aggregated data and statistical analyses. Additionally, we might cite (optional) write-in answers from comment fields.*

Contact person in case of questions about or comments on this study: [Contact data of first author].

Thank you very much for participating!

☐ Ich stimme zu, dass meine Antworten wie oben beschrieben zu Forschungszwecken gespeichert und ausgewertet werden. *I consent to my answers being stored and analyzed for research purposes as outlined above.*

The survey only progresses if this box is checked.

————

In dieser Umfrage untersuchen wir Sprachen und Sprachformen, die sich deutlich vom Hochdeutschen unterscheiden. Damit meinen wir mit dem Deutschen verwandte **Regionalsprachen** sowie **Dialekte, Mundarten** und **Platt**-Varianten, die meist für eine kleine Region typisch sind und von Außenstehenden nicht ohne Weiteres verstanden werden können. Ein paar Beispiele dafür sind das Eifeler Platt, Allgäuerisch, Bairisch oder Nordfriesisch. **Der Einfachheit halber verwenden wir im Folgenden „Dialekt" als Sammelbegriff für all diese Sprachformen.** *In this survey, we focus on languages and language varieties that*

are clearly distinct from Standard German. To be precise, we are interested in **regional languages** *related to German as well as* **dialects**[8] *that usually are typical for a small region and cannot easily be understood by outsiders. Some examples are Eifelplatt, Allgäu dialects, Bavarian and North Frisian.* **For the sake of simplicity, we will use "dialect" as umbrella term for all of these language varieties in the following.**

This introduction is partially based on the one from the REDE project surveys (Schmidt et al., 2020–).

1. Können Sie einen deutschen Dialekt sprechen? *Can you speak a German dialect?*

Ja, sehr sicher / *Yes, very well*	223
Ja, gut / *Yes, well*	71
Ein wenig / *A bit*	33
Nein / *No*	16
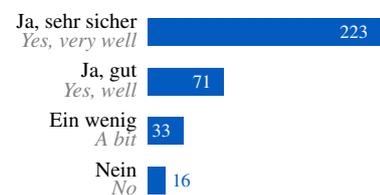
The 16 respondents who answered 'no' are excluded from the analysis. The survey automatically ended for them, showing the message: "Alle weiteren Fragen richten sich nur an SprecherInnen eines deutschen Dialekts bzw. einer mit dem Deutschen nahe verwandten Regionalsprache. Vielen Dank für Ihre Teilnahme!" *"All further questions are only for speakers of a German dialect or a regional language closely related to German. Thank you for participating!"*

2. Um welchen Dialekt handelt es sich? *Which dialect specifically?*

327 write-in answers.

3. Wann haben Sie diesen Dialekt gelernt? *When did you learn this dialect?*

Als Muttersprache / *As mother tongue*	231
Kindheit/Jugend / *Childhood/youth*	84
Später / *Later*	12
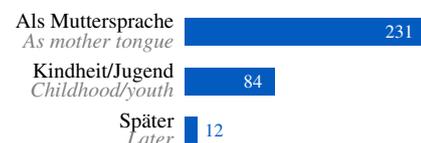
4. In welchem Land befindet sich der Ort, an dem Ihr Dialekt gesprochen wird (z.B. Ihr Heimatort)? *In which country is the location where your dialect is spoken (e.g., your hometown)?*

See Figure 1.

---

[8]*In the German version, we include different terms that all translate to "dialect" but are used in different regions.*

5. In welchem Bundesland befindet sich dieser Ort? *In which German state is this location?*

Only asked if the previous answer is 'Germany'. See Figure 1.

6. Wie sehr entspricht Ihr Dialekt dem traditionellen Dialekt des Ortes? *How much does your dialect resemble the traditional dialect of this location?*

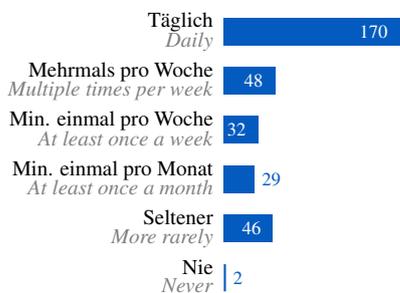|	
1 – Mein Dialekt ist sehr traditionell...*	57
2	114
3	93
4	38
5 – Mein Dialekt ist eher regional gefärbtes...**	25

*1 – Mein Dialekt ist sehr traditionell und für Außenstehende aus anderen Regionen sehr schwer zu verstehen. *1 – My dialect is very traditional and very hard to understand for outsiders from other regions.*

**5 – Mein Dialekt ist eher regional gefärbtes Deutsch, das auch von Außenstehenden recht einfach verstanden wird. *5 – My dialect is more like regionally marked German that is relatively easily understood by outsiders.*

7. Wie häufig sprechen Sie Ihren Dialekt? *How often do you speak your dialect?*

The answer options are based on those in the surveys summarized by Soria et al. (2018).

Täglich / *Daily*	170
Mehrmals pro Woche / *Multiple times per week*	48
Min. einmal pro Woche / *At least once a week*	32
Min. einmal pro Monat / *At least once a month*	29
Seltener / *More rarely*	46
Nie / *Never*	2

8. Schreiben Sie manchmal Ihren Dialekt? *Do you ever write your dialect?*

This question and the next one are modelled after questions by Millour (2020, pp. 228, 237–238).

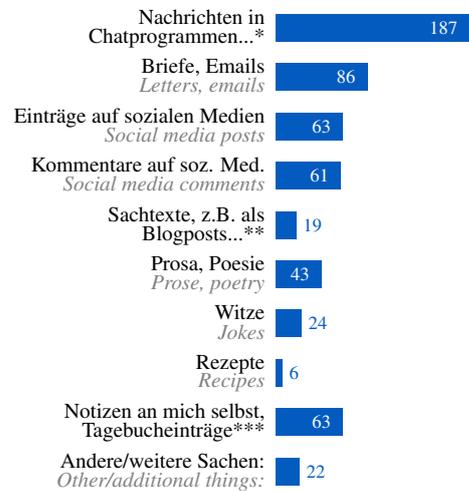|	
Ja (egal ob häufig o. selten) / *Yes (whether often or rarely)*	217
Nein, ich weiß nicht wie / *No, I don't know how*	14
Nein, ich habe dazu keine Gelegenheit*	26
Nein, mein Dialekt ist eine gesprochene Sprachform...**	62
Nein, aus anderen Gründen: / *No, for other reasons:*	8

*No, I don't have any opportunity for this

**Nein, mein Dialekt ist eine gesprochene Sprachform und ich möchte ihn nicht schreiben *No, my dialect is a spoken form of language and I don't want to write it*

9. Was schreiben Sie in Ihrem Dialekt? (Mehrfachantworten möglich) *What do you write in your dialect? (Multiple answers possible)*

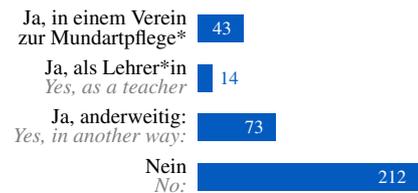*Only asked if previous is 'yes'. 217 participants responded:*

Nachrichten in Chatprogrammen...*	187
Briefe, Emails / *Letters, emails*	86
Einträge auf sozialen Medien / *Social media posts*	63
Kommentare auf soz. Med. / *Social media comments*	61
Sachtexte, z.B. als Blogposts...**	19
Prosa, Poesie / *Prose, poetry*	43
Witze / *Jokes*	24
Rezepte / *Recipes*	6
Notizen an mich selbst, Tagebucheinträge***	63
Andere/weitere Sachen: / *Other/additional things:*	22

*Nachrichten in Chatprogrammen, Messengern (wie WhatsApp), SMS *Texts in messaging apps (like WhatsApp), text messages*

**Sachtexte, z.B. als Blogposts oder auf Wikipedia *Non-fiction texts, e.g., blog posts or Wikipedia articles*
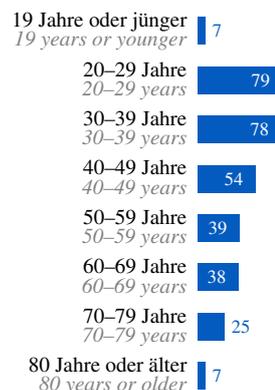
***Notes to myself, diary entries

10. Setzen Sie sich aktiv für den Erhalt Ihres Dialekts ein? (Mehrfachantworten möglich) *Are you actively involved in preserving your dialect? (Multiple answers possible)*

This question is based on questions in the surveys by Soria et al. (2018) and Millour (2020, pp. 227, 235). 323 respondents answered:

Ja, in einem Verein zur Mundartpflege*	43
Ja, als Lehrer*in / *Yes, as a teacher*	14
Ja, anderweitig: / *Yes, in another way:*	73
Nein / *No:*	212

*Yes, in a dialect preservation society

11. Wie alt sind Sie? *How old are you?*

19 Jahre oder jünger / *19 years or younger*	7
20–29 Jahre / *20–29 years*	79
30–39 Jahre / *30–39 years*	78
40–49 Jahre / *40–49 years*	54
50–59 Jahre / *50–59 years*	39
60–69 Jahre / *60–69 years*	38
70–79 Jahre / *70–79 years*	25
80 Jahre oder älter / *80 years or older*	7

12. Weitere Kommentare zu Ihrem Dialekt oder zu den vorherigen Fragen: (Optional) *Additional comments on your dialect or the preceding questions: (Optional)*

61 write-in answers.

———

In diesem Abschnitt fragen wir Sie zu Ihrer Meinung zu verschiedenen dialektbezogenen Themen. **Dabei gibt es keine richtigen/falschen oder erwünschten/unerwünschten Antworten,** sondern wir sind an Ihrer persönlichen Meinung interessiert. *In this section we ask you about your opinion on different dialect-related topics.* **There are no right/wrong or desired/undesired answers,** *we are simply interested in your personal opinion.*

13. Stimmen Sie den folgenden Aussagen zu? *Do you agree with the following statements?*

Statements presented in a randomized order:

- Die Vielfalt der unterschiedlichen Ausprägungen meines Dialekts ist eine Stärke. *The diversity of the different varieties of my dialect is a strength.*

- Mein Dialekt ist in erster Linie eine gesprochene Sprachform und nicht für die schriftliche Kommunikation geeignet. *My dialect is primarily a spoken form of language and not suited for written communication.* **This question is based on an answer option in the survey by** *Millour (2020, pp. 228, 237) (see also question 8 in this appendix).*

- Ich möchte meinen Dialekt in allen Lebensbereichen verwenden. *I'd like to be able to use my dialect in any aspect of life.* **This question is based on a question by** *Soria et al. (2018) and Millour (2020, pp. 229, 239).*

- Wenn ich einen Text lese, den jemand anderes in meinem Dialekt verfasst hat, fällt es mir schwer, ihn zu verstehen. *When I read text that someone else wrote in my dialect, I have trouble understanding it.*

- Es sollte eine standardisierte Rechtschreibung für meinen Dialekt geben. *There should be a standardized orthography for my dialect.*

Answer options:

- Ja, auf jeden Fall *Yes, absolutely*

- Eher ja *Rather yes*

- Weder noch *Neither/nor*

- Eher nein *Rather no*

- Nein, gar nicht *Absolutely not*

- Keine Angabe *Prefer not to say*

The answer distributions (in %) are as follows:
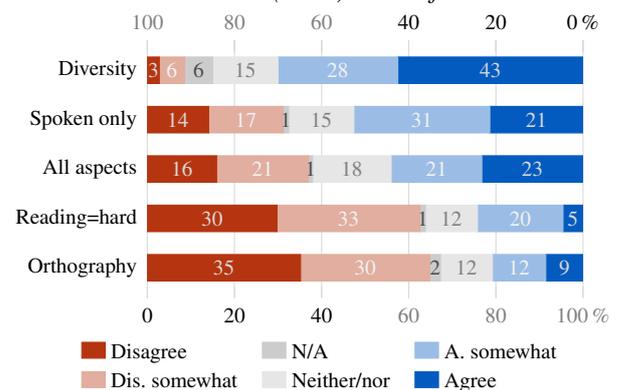
14. Weitere Kommentare zu diesem Abschnitt: (Optional) *Additional comments on this section: (Optional)*

48 write-in answers.

———

In diesem Abschnitt fragen wir Sie zu Ihrer Meinung zu verschiedenen Sprachtechnologien. **Dabei gibt es keine richtigen/falschen oder erwünschten/unerwünschten Antworten,** sondern wir sind an Ihrer persönlichen Meinung interessiert. *In this section we ask you about your opinion on different language technologies.* **There are no right/wrong or desired/undesired answers,** *we are simply interested in your personal opinion.*

> **Übersetzungsprogramme** erstellen eine automatische Übersetzung von Text aus einer Sprache in eine andere Sprache. Beispiele dafür sind DeepL oder Google Translate. *Machine translation software automatically translate text from one language into another. Examples are DeepL or Google Translate.*

15. Stimmen Sie den folgenden Aussagen zu? Es sollte Übersetzungsprogramme geben, ... *Do you agree with the following statements? There should be translation software...*

- ...die hochdeutsche Texte in meinen Dialekt übersetzen. *...that translates Standard German texts into my dialect.*

- ...die Texte aus anderen Sprachen in meinen Dialekt übersetzen. *...that translates texts from other languages into my dialect.*

- ...die Texte aus meinem Dialekt ins Hochdeutsche übersetzen. *...that translates texts from my dialect into Standard German.*

- ...die Texte aus meinem Dialekt in andere Sprachen übersetzen. *...that translates texts from my dialect into other languages.*

Answer options:

- Ja, unbedingt *Yes, absolutely*

- Eher ja *Rather yes*

- Weder noch *Neither/nor*

- Eher nein *Rather no*

- Nein, das halte ich nicht für sinnvoll *No, I don't think this is useful*

- Das kann ich nicht bewerten *I cannot judge this*

See Figure 2 for answer distributions.

16. Welcher Aussage stimmen Sie mehr zu? Wenn ich einen Text in meinen Dialekt übersetzen lasse, ... *With which statement do you agree more? When a text is translated into my dialect, ...*

1 – ...soll die Übersetzung sprachlich meiner...*	52
2	61
3	67
4	32
5 – ...ist es mir egal, welcher (geschriebenen)...**	43
Ich möchte keinen geschriebenen Dialekt-Output.***	72

*1 – ...soll die Übersetzung sprachlich meiner (geschriebenen) Version des Dialekts voll und ganz entsprechen. *1 – ...the translation should fully correspond to my own (written) version of the dialect.*

**5 – ...ist es mir egal, welcher (geschriebenen) Form meines Dialekts die Übersetzung sprachlich entspricht. *5 – ...I do not care which (written) version of my dialect the translation corresponds to.*

***I do not want any written dialect output.*

17. Weitere Kommentare zu Übersetzungsprogrammen: (Optional) *Additional comments on machine translation software: (Optional)*

41 write-in answers.

> **Rechtschreib- und Grammatikkorrekturprogramme** markieren oder korrigieren mögliche Fehler in Texten, zum Beispiel bei der Eingabe in Microsoft Word. ***Spell- and grammar checkers** highlight or fix potential errors in texts, for instance when writing text in Microsoft Word.*

18. Stimmen Sie der folgenden Aussage zu? Es sollte Rechtschreib- und Grammatikkorrekturprogramme für meinen Dialekt geben. *Do you agree with the following statement? There should be spell- and grammar checkers for my dialect.*

Same answer options as for question 15. See Figure 2 for the answer distribution.

19. Weitere Kommentare zu Rechtschreib- und Grammatikkorrekturprogrammen: (Optional) *Additional comments on spell- and grammar checkers: (Optional)*

51 write-in answers.

> **Transkriptionsprogramme** verschriftlichen gesprochene Sprache. Sie finden beispielsweise bei automatisch erzeugten Untertiteln oder bei Diktiergeräten Einsatz. ***Speech-to-text systems** transcribe spoken language. They are for instance used for automatically generating subtitles or in the context of dictation software.*

20. Stimmen Sie den folgenden Aussagen zu? Es sollte Transkriptionsprogramme geben, ... *Do you agree with the following statements? There should be speech-to-text software...*

- ...die Audioaufnahmen in meinem Dialekt als geschriebenes Hochdeutsch wiedergeben. *...that transcribes audio recorded in my dialect as written Standard German.*

- ...die Audioaufnahmen in meinem Dialekt als geschriebenen Dialekt wiedergeben. *...that transcribes audio recorded in my dialect as written dialect.*

21. Weitere Kommentare zu Transkriptionsprogrammen: (Optional) *Additional comments on speech-to-text software: (Optional)*

33 write-in answers.

> **Text-to-Speech-Systeme** funktionieren umgekehrt wie Transkriptionsprogramme: sie erzeugen gesprochene Versionen von geschriebenem Text. Ein Beispiel dafür sind Bildschirmleseprogramme. *Text-to-speech systems work the other way around as speech-to-text systems: they generate spoken versions of written text. One example are screen readers.*

22. Stimmen Sie der folgenden Aussage zu? Es sollte Text-to-Speech-Systeme geben, die meinen Dialekt von geschriebener Form in gesprochene Form umwandeln. *Do you agree with the following statement? There should be text-to-speech systems that synthesize dialectal audio for text written in my dialect.*

23. Weitere Kommentare zu Text-to-Speech-Systemen: (Optional) *Additional comments on text-to-speech systems: (Optional)*
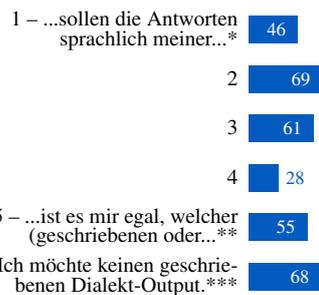
22 write-in answers.

> **Sprachassistenten** sind Programme, die geschriebene oder gesprochene Fragen beantworten bzw. Befehle ausführen, zum Beispiel Siri oder Alexa.
> Eng verwandt damit sind **Chatbots**: Programme, die textbasierte Dialoge ermöglichen, bei denen ein Programm Antworten auf Texteingaben von Nutzer*innen erzeugt. Ein Beispiel dafür ist ChatGPT.
> *Digital assistants are programs that answer written or spoken questions and carry out commands, like Siri or Alexa.*
> *Chatbots are closely related. They are software that enables text-based dialogues, wherein a program generates answers to text input from users. An example is ChatGPT.*

24. Stimmen Sie den folgenden Aussagen zu? *Do you agree with the following statements?*

- Es sollte Sprachassistenten geben, die man mit Fragen/Befehlen in meinem Dialekt bedienen kann. *There should be digital assistants that you can query with questions/commands in my dialect.*

- Es sollte Sprachassistenten geben, die in meinem Dialekt auf Fragen/Befehle antworten. *There should be digital assistants that use my dialect when replying to questions/commands.*

- Es sollte Chatbots geben, die auf Eingaben in meinem Dialekt antworten können. *There should be chatbots that can respond to inputs written in my dialect.*

- Es sollte Chatbots geben, deren Antworten in meinem Dialekt verfasst sind. *There should be chatbots who respond in my dialect.*

25. Welcher Aussage stimmen Sie mehr zu? Wenn ein Sprachassistent oder ein Chatbot Antworten in meinem Dialekt erzeugt, ... *With which statement do you agree more? When a digital assistant or chatbot generates replies in my dialect, ...*

Option	Count
1 – ...sollen die Antworten sprachlich meiner...*	46
2	69
3	61
4	28
5 – ...ist es mir egal, welcher (geschriebenen oder...**	55
Ich möchte keinen geschriebenen Dialekt-Output.***	68

*1 – ...sollen die Antworten sprachlich meiner (geschriebenen oder gesprochenen) Version des Dialekts voll und ganz entsprechen. *1 – ...the replies should fully correspond to my own (written or spoken) version of the dialect.*

**5 – ...ist es mir egal, welcher (geschriebenen oder gesprochenen) Form meines Dialekts die Antworten sprachlich entsprechen. *5 – ...I do not care which (written or spoken) version of my dialect the replies correspond to.*

***I do not want any written dialect output.*

26. Weitere Kommentare zu Sprachassistenten oder Chatbots: (Optional) *Additional comments on digital assistants or chatbots: (Optional)*

25 write-in answers.

27. Stimmen Sie der folgenden Aussage zu? Es sollte Suchmaschinen geben, bei denen ich meinen Dialekt als Eingabesprache verwenden kann. *Do you agree with the following statement? There should be search engines that support queries in my dialect.*

Same answer options as for question 15. See Figure 2 for the answer distribution.

28. Weitere Kommentare zu Suchmaschinen:
(Optional) *Additional comments on search engines: (Optional)*

15 write-in answers.

---

29. Sind Ihnen bereits Sprachtechnologien bekannt, die Ihren Dialekt unterstützen? *Are you already aware of any language technologies for your dialect?*


30. Stimmen Sie den folgenden Aussagen zu?
Do you agree with the following statements?
Statements presented in a randomized order:

- Sprachtechnologie, die ich für sinnvoll halte, nutze ich auch selbst. *If I find language technology useful, I also use it myself.*

- Eine größere Unterstützung durch Sprachtechnologie würde meinen Dialekt attraktiver für jüngere Generationen machen. *If my dialect were supported more by language technologies, the dialect would be more appealing for younger generations. This question is modelled after questions in the surveys by Soria et al. (2018) and Millour (2020, p. 229), asking about the hypothesized impact of a language's increased use online on the appeal for younger people.*

- Wenn ich Sprachtechnologie für meinen Dialekt hätte, würde ich ihn häufiger verwenden. *If I had language technology for my dialect, I would use my dialect more often.*
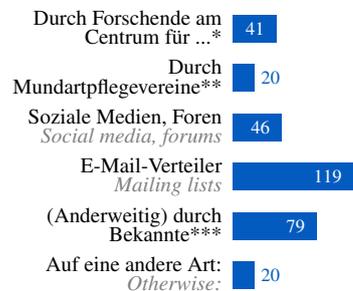
See question 13 for the answer options. Answer distributions (in %):
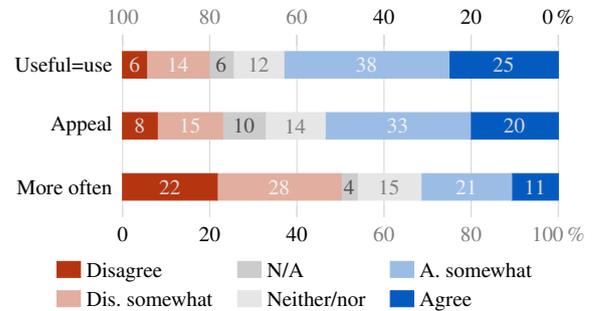
31. Weitere Kommentare zum Thema Sprachtechnologie oder allgemein zu dieser Umfrage:
(Optional) *Additional comments on language technology or generally regarding this survey: (Optional)*

29 write-in answers.

32. Wie haben Sie von dieser Studie erfahren?
How did you find out about this study?

Durch Forschende am Centrum für Informations- und Sprachverarbeitung (LMU) Via researchers at the Center for Information and Language Processing (LMU)
**Via dialect preservation societies*
***(Otherwise) via acquaintances*

Vielen Dank für Ihre Teilnahme! Wir möchten uns ganz herzlich für Ihre Mithilfe bedanken. Ihre Antworten wurden gespeichert, Sie können das Browser-Fenster nun schließen. *Thank you for participating! We would like to thank you very much for your help. Your answers have been saved; you can close the browser window now.*

Rank	All		Non-activists only	
	LTs	Mean	LTs	Mean
1	Assistant in (24)	3.75	Assistant in (24)	3.80
2	STT deu (20)	3.46	STT deu (20)	3.48
3	STT dial (20)	3.38	Chatbot in (24)	3.25
4	Chatbot in (24)	3.29	STT dial (20)	3.24
5	MT dial→deu (15)	3.17	Assistant out (24)	3.01
6	Assistant out (24)	3.14	MT dial→deu (15)	3.00
7	TTS (22)	3.13	TTS (22)	2.99
8	Search engines (27)	2.94	Search engines (27)	2.69
9	Chatbot out (24)	2.76	Chatbot out (24)	2.59
10	MT dial→oth (15)	2.73	MT dial→oth (15)	2.59
11	MT deu→dial (15)	2.71	MT deu→dial (15)	2.53
12	MT oth→dial (15)	2.39	MT oth→dial (15)	2.17
13	Spellcheckers (18)	2.38	Spellcheckers (18)	2.08

Table 1: **Language technologies sorted by mean score** given by all respondents and non-activists only (participants who did not indicate involvement in language preservation, §4.4). 'Mean' refers to the mean Likert score (see text). Numbers behind the LT names refer to questions in §A.

B Correlation Scores

Figure 3 shows the Spearman's rank correlation coefficients (ρ) between the variables investigated in the questionnaire, with ρ values ranging from –0.50 to +0.77.

For the correlation analysis and the subgroup comparisons (Appendix §D), the variable values are mapped so that higher values correspond to higher agreement with the statements in questions 13, 15, 18, 20, 22, 24, 27 and 30, and to higher dialect competence (question 1) and usage frequency (7), higher age (11) and age of dialect acquisition (3), more traditional dialects (6),[9] and greater openness towards variation in the output of MT (16) and digital assistants / chatbots (25). The variable *# writing contexts* encodes the number of answer options selected in question 9. The variables *writing* (8) and *activism* (10) are binary such that 0 encodes the 'no' options and 1 stands for the 'yes' options.

The beginning of the first row in the figure can thus be read as follows: Dialect competence self-ratings are

- negatively correlated with the age of acquisition (i.e., respondents whose dialect is their first language generally give higher competence ratings),

---

[9]Note that this the inverse of how the question is originally phrased.

- slightly positively correlated with language activism (i.e., fluent dialect speakers are slightly more likely to be engaged in dialect preservation activities, and vice versa),

- positively correlated with traditionality (i.e., competent dialect speakers tend to rate their dialect as more distinct from Standard German, and vice versa),

and so on.

C LT Ranking

Table 1 shows the order of preferred LTs. This ranking is based on the mean scores when coding the answers as follows: 1 = useless, 2 = rather useless, 3 = neither/not, 4 = rather useful, 5 = useful. Non-answers ('cannot judge') are excluded.

If we remove the participants who indicated active engagement in dialect preservation (see §4.4 and question 10), the ranking only changes very slightly: chatbots with dialectal input and STT with dialectal output trade places (although they have nearly identical mean scores), and we observe the same for virtual assistants with dialectal output and machine translation from the dialect into Standard German.

D Subgroup Comparisons

Tables 2 and 3 show how the responses by different subgroups of respondents differ for each variable.

We provide each subgroup's mean answers (using the same numeric coding as in the previous two appendix sections), as well as *t*-test statistics (taking into account the scalar nature of the answer options) and χ^2 test results.

In addition to the analyses in §4.4, we provide two more subgroup comparisons, albeit with small effect sizes:

Traditionality   We compare the responses of speakers who rate their dialect as traditional and distinct from Standard German (the first two answer options for question 6) to those who indicated speaking a variety more akin to a regiolect (the last two options). While these subgroups differ in their responses to the dialect-related questions, few of the differences regarding language technologies are statistically significant (Table 2).

Age   Figure 3 shows that the variable *age* correlates with few other variables. With respect to the LTs, young participants tend to be somewhat more positive towards three of the overall most popular LTs: STT with Standard German output, and virtual assistants and chatbots with dialectal input.

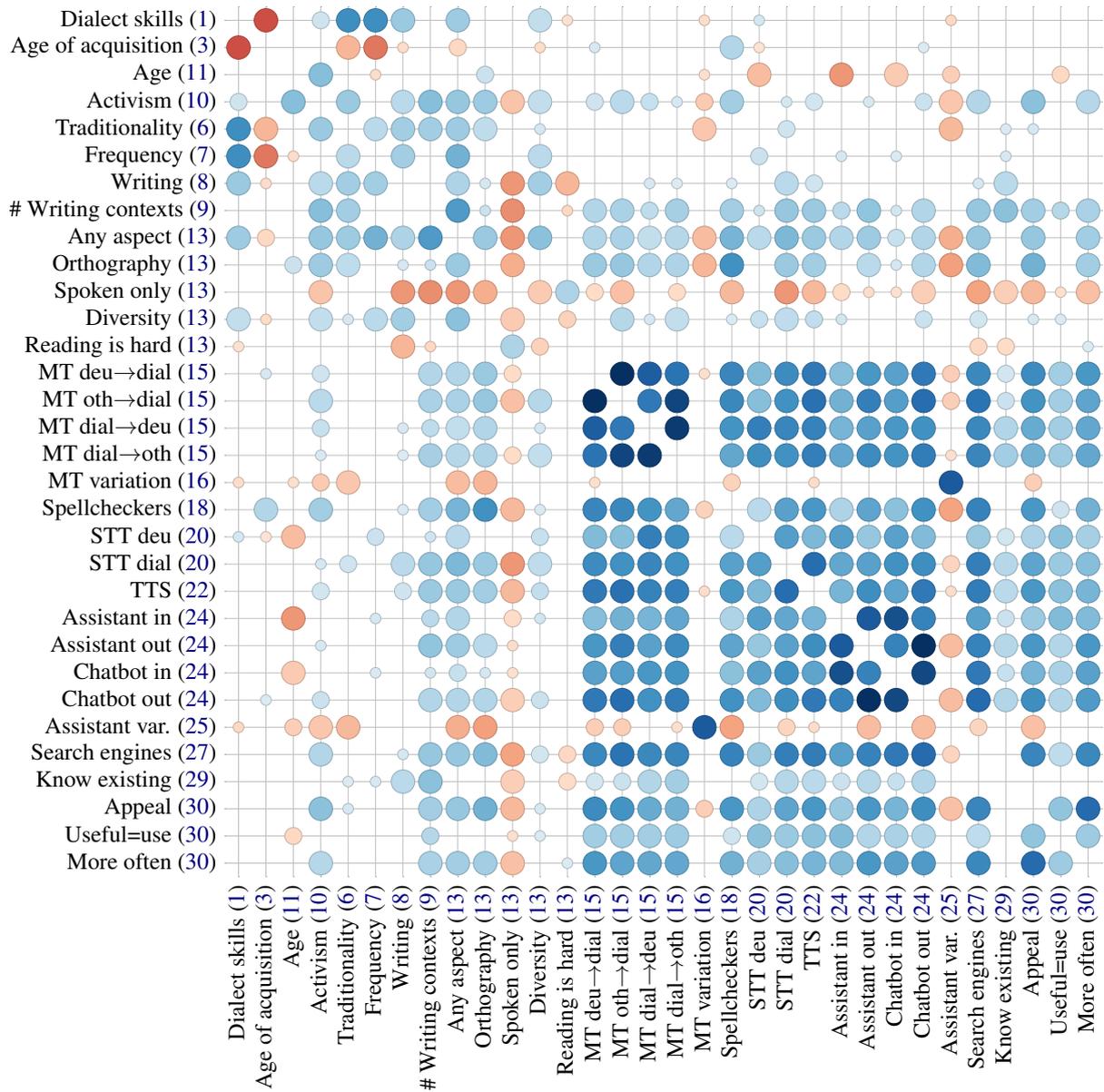Figure 3: **Spearman's ρ between variables.** Blue dots show positively correlated variables (max.: +0.77), red dots negatively correlated ones (min.: -0.50). We only include correlations with *p*-values under 0.05. The larger the dot, the smaller the *p*-value. The numbers behind the variables refer to the questions in Appendix §A. For further explanations of how the variables are coded, see Appendix §B.
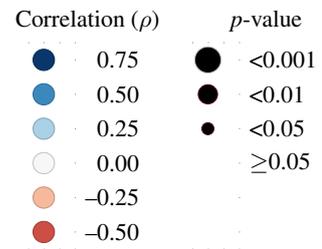
Variable	Activists vs. non-activists				Most vs. least trad. dialects			
	t-stat	χ^2	μ_{Act}	μ_{Non}	t-stat	χ^2	μ_{Most}	μ_{Least}
Dialect skills (1)	3.0 **	9.0 *	3.7	3.5	11.3 ***	83.2 ***	3.8	2.9
Age of acquisition (3)	1.3	5.4	1.4	1.3	-4.4 ***	29.9 ***	1.2	1.5
Age (11)	6.7 ***	43.5 ***	4.7	3.4	0.7	3.9	3.9	3.7
Activism (10)	—	—	1.0	0.0	5.7 ***	26.9 ***	0.5	0.1
Traditionality (6)	5.6 ***	31.6 ***	3.9	3.2	—	—	4.3	1.6
Frequency (7)	2.5 *	14.0 *	5.1	4.6	5.0 ***	25.9 ***	5.1	4.0
Writing (8)	3.9 ***	13.9 ***	0.8	0.6	5.0 ***	21.3 ***	0.8	0.4
# Writing contexts (9)	4.9 ***	28.7 ***	3.3	2.2	3.4 ***	16.2 *	3.0	1.8
Any aspect (13)	5.6 ***	36.3 ***	3.7	2.8	6.2 ***	36.4 ***	3.5	2.2
Orthography (13)	6.0 ***	37.9 ***	2.8	2.0	2.8 **	12.6 *	2.5	1.9
Spoken only (13)	-4.2 ***	34.9 ***	2.9	3.5	-2.5 *	8.7	3.1	3.7
Diversity (13)	3.0 **	12.5 *	4.3	3.9	3.0 **	12.5 *	4.2	3.7
Reading is hard (13)	-1.6	5.5	2.2	2.4	-2.3 *	10.4 *	2.3	2.7
MT deu→dial (15)	3.0 **	14.7 **	3.0	2.5	0.9	2.0	2.7	2.5
MT oth→dial (15)	4.1 ***	21.8 ***	2.8	2.2	1.6	12.3 *	2.5	2.2
MT dial→deu (15)	3.1 **	13.1 *	3.5	3.0	1.5	7.9	3.2	2.9
MT dial→oth (15)	2.4 *	7.0	3.0	2.6	2.0 *	20.0 ***	2.9	2.4
MT variation (16)	-3.3 **	11.7 *	2.5	3.0	-4.2 ***	19.1 ***	2.6	3.6
Spellcheckers (18)	5.4 ***	34.2 ***	2.9	2.1	2.0 *	9.8 *	2.4	2.0
STT deu (20)	-0.3	3.1	3.4	3.5	1.0	8.9	3.5	3.3
STT dial (20)	2.4 *	8.7	3.6	3.2	2.6 **	10.1 *	3.5	2.9
TTS (22)	2.5 *	11.6 *	3.4	3.0	1.0	6.9	3.1	2.9
Assistant in (24)	-0.8	3.3	3.7	3.8	1.0	1.6	3.8	3.6
Assistant out (24)	2.3 *	5.6	3.4	3.0	-0.1	1.7	3.1	3.1
Chatbot in (24)	0.7	1.3	3.4	3.2	0.2	3.7	3.3	3.3
Chatbot out (24)	3.1 **	12.1 *	3.1	2.6	0.6	2.0	2.8	2.7
Assistant var. (25)	-3.5 ***	13.7 **	2.5	3.1	-4.9 ***	22.9 ***	2.7	3.7
Search engines (27)	4.2 ***	20.8 ***	3.4	2.7	0.9	4.2	2.9	2.7
Know existing (29)	0.6	0.2	0.2	0.1	2.0 *	3.1	0.2	0.1
Appeal (30)	5.4 ***	32.1 ***	4.0	3.2	1.5	8.0	3.6	3.3
Useful=use (30)	0.5	5.0	3.7	3.6	1.3	3.4	3.7	3.5
More often (30)	4.0 ***	16.5 **	3.1	2.5	0.1	1.0	2.6	2.6

Table 2: **Differences between respondent subgroups.** We show the results of t-tests and χ^2 tests between pairs of respondent groups: those who indicated involvement in dialect preservation efforts ('activists', question 10) vs. those who did not, and respondents who rate their dialect as one of the two most vs. two least traditional options (question 6). Positive t-statistics indicate that the first group's values for the variable are higher than the second one's, and vice versa for negative values. Grey entries denote results with p-values ≥ 0.05; asterisks represent smaller p-values: $* < 0.05$, $** < 0.01$, $*** < 0.001$. The columns with μ present the mean Likert scores of the subgroups' responses (e.g., μ_{Act} contains the activists' mean answers). The numbers behind the variables refer to the questions in Appendix §A. For information on the variables on how the variables are encoded as numbers, see Appendix §B.

	NDS vs. (rest of) D/AT		CH vs. (non-NDS) D/AT				
	t-stat	χ^2	t-stat	χ^2	μ_{NDS}	$\mu_{\text{D/AT}}$	μ_{CH}
Dialect skills (1)	-1.6	5.5	4.7 ***	23.4 ***	3.4	3.5	4.0
Age of acquisition (3)	5.2 ***	25.2 ***	-3.2 **	10.0 **	1.7	1.3	1.1
Age (11)	6.3 ***	38.2 ***	-1.5	15.1 *	5.3	3.6	3.2
Activism (10)	5.3 ***	23.9 ***	-1.8	2.5	0.7	0.3	0.2
Traditionality (6)	2.8 **	9.9 *	3.3 **	16.6 **	3.7	3.3	3.8
Frequency (7)	-1.1	6.4	4.6 ***	23.0 ***	4.4	4.7	5.7
Writing (8)	1.2	1.1	3.8 ***	12.5 ***	0.7	0.6	0.9
# Writing contexts (9)	6.5 ***	45.9 ***	4.5 ***	25.5 ***	4.0	2.1	3.2
Any aspect (13)	5.3 ***	27.6 ***	3.9 ***	27.3 ***	3.9	2.8	3.7
Orthography (13)	8.4 ***	63.6 ***	-2.0 *	4.8	3.6	2.0	1.7
Spoken only (13)	-5.6 ***	30.6 ***	-3.0 **	19.7 ***	2.5	3.6	3.0
Diversity (13)	1.0	3.1	2.7 **	7.3	4.2	4.0	4.5
Reading is hard (13)	-4.1 ***	18.3 **	-1.8	3.2	1.8	2.6	2.2
MT deu→dial (15)	4.6 ***	21.3 ***	-0.5	1.6	3.5	2.5	2.4
MT oth→dial (15)	5.0 ***	27.8 ***	-0.2	5.0	3.2	2.2	2.2
MT dial→deu (15)	2.6 **	9.9 *	1.0	4.3	3.6	3.0	3.3
MT dial→oth (15)	3.3 **	13.6 **	1.3	4.4	3.2	2.5	2.8
MT variation (16)	-2.4 *	11.6 *	0.7	4.5	2.3	2.9	3.0
Spellcheckers (18)	8.2 ***	68.3 ***	-2.1 *	8.2	3.7	2.1	1.7
STT deu (20)	-1.5	9.5	3.2 **	10.5 *	3.1	3.4	4.1
STT dial (20)	4.0 ***	17.5 **	2.6 **	10.7 *	4.0	3.1	3.7
TTS (22)	4.0 ***	15.9 **	0.5	1.5	3.8	2.9	3.1
Assistant in (24)	-0.8	7.6	2.7 **	9.1	3.5	3.7	4.2
Assistant out (24)	1.7	3.0	0.8	2.6	3.4	3.0	3.2
Chatbot in (24)	0.8	2.0	-0.6	2.1	3.4	3.2	3.1
Chatbot out (24)	3.7 ***	13.4 **	-0.4	2.8	3.4	2.6	2.5
Assistant var. (25)	-2.3 *	9.2	1.1	2.1	2.4	2.9	3.2
Search engines (27)	4.8 ***	26.1 ***	-1.6	3.8	3.8	2.8	2.4
Know existing (29)	4.1 ***	13.6 ***	8.2 ***	49.5 ***	0.2	0.1	0.5
Appeal (30)	5.1 ***	35.3 ***	-3.1 **	14.0 **	4.3	3.3	2.7
Useful=use (30)	0.1	1.4	1.6	3.8	3.6	3.6	3.9
More often (30)	3.7 ***	14.8 **	-1.6	3.1	3.2	2.5	2.2

Table 3: **Differences between region-based respondent subgroups.** We show the results of t-tests and χ^2 tests between Low German (NDS) or Swiss (CH) respondents compared to (non-Low-German-speaking) German and Austrian respondents (D/AT). Positive t-statistics indicate that the first group's values for the variable are higher than the second one's, and vice versa for negative values. Grey entries denote results with p-values ≥ 0.05; asterisks represent smaller p-values: $* < 0.05$, $** < 0.01$, $*** < 0.001$. The columns with μ present the mean Likert scores of the subgroups' responses (e.g., μ_{NDS} contains the mean answers provided by our Low Saxon respondents). The numbers behind the variables refer to the questions in Appendix §A. For information on the variables on how the variables are encoded as numbers, see Appendix §B.