The Influence of Intrinsic and Extrinsic Motivation on the Creation of Language Resources in a Citizen Linguistics Project about Lexicography

Barbara Heinisch

Centre for Translation Studies, University of Vienna, Austria
Porzellangasse 4, 1090 Vienna
barbara.heinisch@univie.ac.at

Abstract

In the field of citizen linguistics, various initiatives are aimed at the creation of language resources by members of the public. To recruit and retain these participants different incentives informed by different motivations, extrinsic and intrinsic ones, play a role at different project stages. Illustrated by a project in the field of lexicography which draws on the extrinsic and/or intrinsic motivation of participants, the complexity of providing the 'right' incentives is addressed. This complexity does not only surface when considering cultural differences and the heterogeneity of the motivations participants might have but also through the changing motivations over time. Here, identifying target groups may help to guide recruitment, retention and dissemination activities. In addition, continuous adaptations may be required during the course of the project to strike a balance between necessary and feasible incentives.

Keywords: Language varieties, citizen science, language resource development, German in Austria

1. Introduction

Data collection from people can take many forms, including the generation, compilation or annotation of language resources by 'the crowd'. The reasons for collecting data from people are manifold, including a lack of available resources or the need for authentic data. The incentives given in these projects range from monetary compensation to encourage members of the public to participate in the creation of language resources to other forms of recognizing the participants' contributions.

In the case of the Austrian variety of the German language, the diversity of the language resources available is still rather low, partially due to non-availability and the low number of speakers, e.g. of regional dialects, partially also due to the variability of the data formats and lack of metadata. Although the Austrian Language Resource Portal (Heinisch and Lušicky, 2020) is a first step to coordinate and to pool language resources produced in Austria, it nevertheless has a strong focus on administrative language. Therefore, it does not reflect the diversity of language resources available, and the linguistic diversity of the German language used in Austria.

Not many language resources in Austria cover language varieties, such as dialects (European Language Resource Coordination, 2019; Hegele et al., 2022). Moreover, language resources in Austrian dialects are either non-existent or of small size, thus representing the small number of speakers of a certain dialect. However, language varieties in general, and dialects in particular, are interesting not only to researchers but also to language technology providers (Zampieri et al., 2020), e.g. those specialized in speech-to-text or speech recognition technologies.

In this regard, several citizen science projects in the field of linguistics tried to fill this gap by collecting communication data from people directly. Some initiatives also aim at collecting dialectal data, e.g. What's Up, Switzerland? from the users of chat communication systems (Schweiz forscht, 2021). In these projects, citizen science is usually characterized by unpaid labor. Therefore, the activity of crowdsourcing the generation, preparation and processing of language resources may benefit from the insights gained in the field of citizen science. There is an increasing literature that addresses the motivation of members of the public who contribute to citizen science projects (Moczek, 2019; Raddick et al., 2010).

Therefore, this paper summarizes the different approaches that were used in a linguistic citizen science project aimed at the generation of lexicographical data by members of the public in Austria.

2. Case study: 'Dictionary' creation by citizens

In the following, the case study of the citizen linguistics project 'On everyone's mind and lips – German in Austria' (abbreviated as IamDiÖ in German) is providing an insight into the different incentives employed throughout the project to create and gather language resources in the form of lexicographical entries. Although the resulting 'dictionary' is a language resource of small scale, the characteristic of the language resources developed by 'On everyone's mind and lips – German in Austria' is that they are intended to meet both the researchers' and the participants' needs.

2.1 'Layperson dictionaries'

This project builds on the numerous enterprises of speakers of dialects in Austria to collect and preserve their dialect in form of (online) dictionaries. From a research perspective, 'layperson' dictionaries may not meet the quality criteria for dictionary entries created according to (standardized) lexicographical principles. On the one hand, this makes the data difficult to find and hampers their access, interoperability and re-use (as required by the FAIR principles). On the other hand, since there are no standards that specify orthography for dialects in Austria, for

researchers it is also interesting to investigate how people actually write dialect, either with the standard alphabet or with diacritical or phonetic symbols adapted to their needs.

2.2 Needs of researchers and participants

From the researchers' perspective, the aims of IamDiÖ were to increase the accessibility of language resources that reflect the linguistic diversity, namely the different varieties of the German language used in Austria. For this purpose, participants were invited to collect lexemes in different language varieties in German used in Austria. Therefore, IamDiÖ intended to bridge these so-called 'layperson dictionaries' and dictionaries that were created by professional lexicographers.

For this purpose, the participants were familiarized with the basics of lexicography, i.e. especially the compilation and use of dictionaries with a focus on creating lexicographical entries. This should further support participants in acquiring a critical understanding of their language use and attitudes towards German in Austria.

2.3 The community and its prior experience

Based on a literature review and our previous experience in citizen science projects, different strategies were employed to recruit participants. In the beginning the project primarily targeted people interested in the topic and relied on the already established community of citizen linguists who have already contributed to other activities in the project. Members of the IamDiÖ community have already completed at least one of the following tasks: a) the Question of the Month that required participants to come up with their (research) question regarding the topic of German language in Austria. Ideally, they would also engage in searching for an answer to their question with the help of researchers (this is the co-creation strand of the project), b) linguistic treasure hunts to study the Austrian linguistic landscape or c) they created memes with a focus on (regional) dialects in Austria (Heinisch, 2020). Some participants also filled d) comic strips. They were provided with a comic strip consisting of three pictures, in which people of different age and gender talked to each other. Since the speech bubbles in the comic strip were empty, the participants were asked to fill them. To illicit linguistic diversity, they were prompted to complete the bubbles with the language or language variety of their choice. We also told them that correct spelling does not matter. This was especially important since dialects of the German language in Austria are only transmitted orally and do not follow a written standard. Moreover, this should avoid embarrassment if the spelling was not perceived as correct. Following the filling of the speech bubbles, the participants were asked to place their comic strip on a standard-non-standard and closeness-distance coordinate system. The underlying purpose of this exercise was to test the hypothesis that people tend to speak non-standard language, especially dialects with people to whom they have a close(r) relationship and use standard language when talking to people to whom the relationship is more distant.

Here, the motivation to engage in this activity was entertainment and fun, especially for children. However, some children also voiced complaints since they have to do similar things in school, which clearly diminished their motivation to engage in 'school activities' in their leisure time.

2.4 Incentives

Over the course of the project, IamDiÖ used different incentives to recruit participants, retain them or reward them for their contributions, including those targeted at the extrinsic motivation and those focusing on the intrinsic motivation of the participants.

2.4.1 Extrinsic motivation: Citizen Science Award

The Citizen Science Award is an initiative in Austria that encourages researchers running a citizen science project to take part in a nation-wide competition. However, the competition is not between citizen science projects but among the participants within a citizen science project. The competition is organized and coordinated by the OeAD, which is Austria's Agency for Education and Internationalisation, which also invites researchers to be part of the Citizen Science Award and selects the projects for the competition. The selected citizen science projects specify a task for the competition, i.e. how participants can contribute to their project. The researchers also have to define criteria for the evaluation of the participants' contributions in the competition. There are two participant categories, either adult individuals or school classes. The winning adults receive a material prize while winning school classes receive money for school activities (OeAD, 2022). Furthermore, the participating projects were also supported by the OeAD throughout the Citizen Science Award period to attract persons and school classes to participate in the competition.

Since IamDiÖ also participated in the Citizen Science Award 2021, the incentives were already provided by the organizers and clearly targeted at extrinsic motivation, i.e. being the winner and receiving a prize. Therefore, it was important to clearly specify the criteria that were applied for determining the winner in each participating citizen science project. In case of IamDiÖ, the participants were required to create lexicographical entries in the project's 'online dictionary' tool *Wortgut*. To win this competition, the quantity of different lexicographical entries and the quantity of the data provided within each lexicographical entry were key.

The focus was on the quantity of the data provided by the participants since quality control and validation of the provided data was not possible. The wide range of lexemes entered in the tool and the high number of dialects in Austria would have necessitated an expert for each and every dialect and for youth language. Since it was not feasible to check the lexicographical data entered by the participants, IamDiÖ relied on the good intentions of the users. However, a 'reporting button' was provided so that users could report discriminating or offensive content published by other users. Furthermore, users could also

label their own entries as being offensive or as containing profanity.

To emphasize the research aspect in the project, school classes were also asked to submit a research report in which they describe their topic of the lexemes collected, the reason for selecting this topic, their motivation as well as their approach. This research report should help the pupils to reflect on their language use on the one hand and to assess the quality of their lexicographical entries, on the other. This reflects again the main maxim of the project that not only the researchers should benefit from public engagement through the provision of language data and resources but also the participants themselves. While IamDiÖ did not conduct empirical studies to investigate the benefits for the participants, the participants' self-reported benefits were to express themselves and to have fun.

The experience gained during the Citizen Science Award showed that it was possible to reach a large number of people who contribute data to a citizen linguistics project to create language resources. Furthermore, the output was promising. However, the participation in the Citizen Science Award required a lot of preparation and intense communication throughout the competition. Especially at the beginning of the competition, the project team organized workshops and consultation hours to familiarize the participants with both the project itself and the criteria of assessment used in the competition. Moreover, the tool used for the creation of lexicographical entries did not only require the provision of help pages and tutorials but also constant technical support.

During the Citizen Science Award, the participants had only a limited amount of time, namely a couple of months to contribute to the different citizen science projects. After the end of the competition, the projects evaluated the participants' contributions and announced the winner(s) in the two categories 'school classes' and 'adult individuals'. Finally, the engagement of the participating schools and adults in the Citizen Science Award initiative is acknowledged and honored in the form of cash and material prizes during a concluding event. As part of a festive ceremony, the winners receive their awards from the leader of the relevant citizen science project and representatives from the OeAD and the Austrian Ministry of Education, Science and Research.

To sum up, the incentives that are relevant in the Citizen Science Award are therefore a competition and the chance of winning a prize (experiencing appreciation and recognition), having personal contact with researchers (this was mentioned by both participant groups, i.e. teachers and adults) and contributing to the advancement of knowledge in academia. This demonstrates that different motivations may play a role at the same time.

Therefore, the incentives focusing on the intrinsic motivation are addressed in the following.

2.4.2   Intrinsic motivation

The collection of lexicographical entries by members of the public also draws on intrinsic motivation. For example, members of the public interested in language can follow their personal interests when engaging in a citizen linguistics project and can become part of a community. Some participants enter data in the 'online dictionary' to preserve their language (preservation efforts) or to make their language use visible as a means of self-expression and expression of identity. Other factors that play a role are the personal contact with the researchers (which may also have an element of receiving appreciation and prestige), gaining an insight into academia or getting access to resources, infrastructure or knowledge that they would not have otherwise. However, also contributing to a greater good, such as the advancement of knowledge or instigating change can play a role.

Nevertheless, for participant retention some form of appreciation, entertainment or novelty may also be necessary. This can take the form of the 'contributor of the month', who can be any participant that contributed significantly to the project in the previous project month. This does not necessarily have to be the person who contributed the most data. It can also be a person who brought the project to the attention of the mayor, who organized an event or who recruited other participants. Any other form of appreciation, such as small gifts, e.g. presents related to the topic of the project or language resource may also provide an incentive to support a project (also) beyond data collection.

2.5   The resulting language resource

After the Citizen Science Award, the language resource to which the participants contributed contained 2,638 lexicographical entries. Since the users can select between distinct levels of difficulty when entering their data, the completeness of the entries varies. Since the entries are neither moderated nor curated, the language resource is currently only available upon request. However, in the future, after the introduction of data validation steps and after collecting additional data, these will become openly accessible.

3.   Discussion

Data collection from people can benefit from the insights gained in citizen science projects since accompanying social science research in citizen science projects sheds light onto the motivations underlying the participation of members of the public in citizen science projects.

Since data collection from people is usually mediated through technology, the aspect of the usability of the system through which participants contribute to a project should not be underestimated. Therefore, in the case of online citizen science, in addition to participant motivation, the usability of technology is of the utmost importance (Nov *et al.*, 2011).

3.1   Target groups

In the IamDiÖ project, we had different target groups, including, broadly speaking, schoolchildren, adolescents, schoolteachers, university students, adults and the elderly having an interest in language. We also learned that not only the needs between these target groups but also within these groups varied significantly. Moreover, the type and the context of participation differed significantly. In the

case of schoolchildren, they may either take part as a school class or individually. This may lead to competition or cooperation either with other students in the class or with other schools.

Identifying and specifying target groups may help to organize and structure the recruitment, retention and dissemination activities in a citizen linguistics project, but the large heterogeneity within the different target groups and the wide range of contexts in which participants may engage in the creation, annotation or processing of language resources cannot be fully anticipated by the researchers during the initial project design phase. Therefore, continuous adaptations may be required during the course of the project itself and it may also be necessary to find a balance between what is necessary and what is feasible.

3.2 Types of incentives

The motivations of the participants in citizen linguistics projects aimed at the creation of language resources may be different and motivations change over time. Therefore, Rotman *et al.* (2014) analyzed the cycle of engagement as well as the associated motivational pivotal points in citizen science projects. According to them, egoism is the initial motivation of both participants and researchers. While participants may want to broaden their horizons and complete an enjoyable activity, researchers cooperate with members of the public to collect large amounts of data or data that would be hard to obtain otherwise. Later in the project, e.g. if a task is completed or a project has ended, the participants reassess their contributions to the project based on their previous experience. In this phase, the initially prevailing motivational factor of egoism may be replaced by collectivism or altruism (Rotman *et al.*, 2012). As the motivation of participants changes over the course of a project, it is crucial to find the appropriate incentives for different stages in the project to sustain participants.

According to Rotman *et al.* (2014), the motivations underlying initial participation in ecological citizen science projects were a) personal interest, such as leisure or hobbies, b) self-promotion, such as social advancement, career progress or reputation, c) self-efficacy, e.g. having an impact on academia and be known to academic researchers in the community as well as d) social responsibility that may be conservation or being proud of one's nation or region. For participant retention, on the other hand, clear shared goals of the academics and the participants are important. Moreover, participants need confirmation that the researchers value their contributions. Other factors are acknowledgement, e.g. in academic papers, or mentorship as well as societal impact in the broad sense, including policy action (Rotman *et al.*, 2014). Therefore, according to Palacin *et al.* (2020), different incentive mechanisms may be used: These can be remuneration, such as micropayments, reputation mechanisms or gamification on the one hand, and non-monetary incentives, such as hedonism-enhancing aspects or social reward, on the other. For initial participation, values that characterize openness to change are crucial. However, in the later project stages and for participant retention, self-transcendence values come into play. The reason for this is that "when extrinsic motivators are self-directed, people will not only perform tasks willingly and enthusiastically but also in a sustained manner" (Palacin *et al.* 2020, 15). To sum up, a focus on rewards as a means of incentives can foster self-enhancement values and therefore lead to a stronger focus on the person itself, and not the project topic at hand. Therefore, during later project stages, retention can be increased by creating ownership of the project among the participants. This can take the form of transparent processes and control by the members of the public.

This is in line with other studies that suggest that extrinsic motivation can help to recruit participants but does fail to retain participants in the long term. Therefore, intrinsic motivation can increase the long-term engagement of participants in a project, for example through providing "experiences of relatedness, capacity building, positive feedback and adapted participation modes" (Triago *et al.*, 2017).

Fischer *et al.* (2021) developed the Nibble-and-Drop Framework to address typical issues regarding recruitment and retention. They differentiate between several types of contributors according to the level of contribution to a project and the duration of participation. Based on five degrees of participation, they categorize participants into 1) initial droppers who sign up to a project but never contribute, 2) nibblers who contribute to a small extent to a project, 3) nibble droppers who contributed to the project before withdrawing from the project after a short period of time, 4) hooked participants who contribute significantly over a longer period of time and 5) hooked droppers who contributed significantly to a project but dropped out after some time. This framework demonstrates that, in some cases, researchers cannot influence the motivation by providing incentives. For example, participants may drop out after they have fulfilled their concrete personal aims. Others might never really commit to a citizen science project but may just engage in 'window shopping' to test whether the project appeals to them.

In addition to all these different motivational factors throughout a citizen science project, also participant demographics and cultural differences play a role. However, addressing these would be beyond the scope of this paper.

3.3 One-time versus regular contributions and superficial versus in-depth contributions

Furthermore, the incentives provided strongly depend on the project's objectives. If a large number of people should be addressed and one-time contributions are sufficient, the incentives may be different from those projects that require repeated contributions of sufficient quality.

Therefore, researchers may differentiate between superficial or in-depth contributions. If a project mainly relies on superficial contributions, i.e. contributions that do not require subject-related, project-related knowledge or knowledge of the academic process and that can be easily completed without providing in-depth explanations and instructions on how to complete the task, then

'crowdsourcing', including monetary incentives and entertainment may suffice. Anonymity may help, i.e. either the participants can stay anonymous, and they do not need to be part of a community and/or the researchers do not engage in personal dialogue with the participants. In this case, the researchers just provide information, e.g. on the website and recruit participants via different platforms, including social media, traditional media or crowdsourcing platforms.

On the other hand, if in-depth contributions are required either by the academics or the participants, then the provision of training and a sense of community may be needed. A community may also result in the necessary (peer) pressure to continue contributing to the project.

While competitions and contests may increase the engagement of some superficial contributors, it may also decrease the motivation of other contributors if they are falling behind too much and are being listed on the bottom of the contributors' list. Therefore, the incentives should also be adjusted to these aspects.

Other aspects that deserve consideration are the mode of participation, which can be online or on site, e.g. in the field as well as the type of communication between researchers and the participants, which can also take place fully online or face-to-face. According to Cappa *et al.* (2020), also online-mediated citizen science projects can benefit from face-to-face interactions between researchers and participants since these can enhance participant motivation. This effect of face-to-face interactions in online-mediated citizen science is most pronounced in older participants.

4. Conclusion

Language resources unfold their potential if they can be re-used. Therefore, the access to, processing and availability of language resources is of major importance to researchers. To get access to or even create language resources, researchers are taking different routes, among these is the engagement of members of the public in the creation or processing of language resources. While data collection from people can take many forms, researchers can get inspiration from citizen science initiatives, also beyond those that are focusing on the creation of language resources when engaging members of the public in collecting language data.

Since the participants in projects aimed at the creation or processing of language resources may be heterogeneous and doing a target group analysis (at different times of the project) to analyze their needs may not be feasible in research projects that are often characterized by limited funding, short project duration and low (personnel) resources for communication.

Therefore, to find a balance between the desired aims and the feasible aims, a small number of different incentives may be offered (from the very beginning or at different project phases) to address different motivations. Some incentives might be targeted at the participants' intrinsic motivation, such as recognition or acknowledgement, while other incentives might address extrinsic motivation, such as competition or monetary compensation.

To conclude, each project may define success and active participation differently. Therefore, there is no one-size-fits-all incentive for participants in citizen linguistics projects. The selection of incentives depends on the project and its objectives. Nevertheless, incentives that focus on extrinsic motivation can help to recruit participants and incentives that target intrinsic motivation can help to retain participants for a longer period of time. Depending on the aim of the language resource project that involves members of the public also the expression of one's identity through language and making their 'speech' heard may be important aspects to be considered when providing incentives.

5. Acknowledgements

6. Bibliographical References

Cappa, F., Laut, J., Nov, O., Giustiniano, L. and Porfiri, M. (2016) Activating social strategies: Face-to-face interaction in technology-mediated citizen science. *Journal of environmental management*, 182, 2016: 10.1016/j.jenvman.2016.07.092.

European Language Resource Coordination (2019) ELRC White Paper. Why Language Data Matters. http://www.lr-coordination.eu/sites/default/files/ELRC_Conference/ELRCWhitePaper.pdf.

Fischer, H., Cho, H. and Storksdieck, M. (2021) Going Beyond Hooked Participants: The Nibble-and-Drop Framework for Classifying Citizen Science Participation. *Citizen Science: Theory and Practice*, 6, 2021: 10.5334/cstp.350.

Hegele, S., Heinisch, B., Popp, A., Marheinecke, K., Rios, A., Gromann, D. *et al.* (2022) European Language Equality. https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D1_16__Language_Report_German_.pdf.

Heinisch, B. (2020) Developing Language Resources with Citizen Linguistics in Austria – A Case Study. In Fiumara, J., Cieri, C., Liberman, M. and Callison-Burch, C. (eds), Citizen Linguistics in Language Resource Development (CLLRD 2020) Proceedings. European Language Resources Association (ELRA), pp. 7–14.

Heinisch, B. and Lušicky, V. (2020) The Austrian Language Resource Portal for the Use and Provision of Language Resources in a Language Variety by Public Administration – a Showcase for Collaboration between Public Administration and a University. In Samy, D., Pérez-Fernández, D. and Arenas-García, J. (eds), 1ˢᵗ Workshop on Language Technologies for Government and Public Administration (LT4Gov) Proceedings: LREC 2020 Workshop Language Resources and Evaluation Conference 11–16 May 2020. European Language Resources Association (ELRA), pp. 28–31.

Moczek, N. (2019) Freiwilliges Engagement für Citizen Science-Projekte im Naturschutz: Konstruktion und Validierung eines Skalensystems zur Messung motivationaler und organisationaler Funktionen, 1. Auflage. Lengerich, Pabst Science Publishers.

Nov, O., Arazy, O. and Anderson, D. (2011) Technology-Mediated Citizen Science Participation: A Motivational Model. In, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.

OeAD (2022) Become a Citizen Science Award project. https://zentrumfuercitizenscience.at/en/for-researchers/become-a-citizen-science-award-project.

Palacin, V., Gilbert, S., Orchard, S., Eaton, A., Ferrario, M. A. and Happonen, A. (2020) Drivers of Participation in Digital Citizen Science: Case Studies on Järviwiki and Safecast. *Citizen Science: Theory and Practice*, 5, 2020: 10.5334/cstp.290.

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K. *et al.* (2010) Galaxy Zoo. Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review*, 9, 2010: 10.3847/AER2009036.

Rotman, D., Hammock, J., Preece, J. J., Boston, C. L., Hansen, D. L., Bowser, A. *et al.* (2014) Does motivation in citizen science change with time and culture? In Fussell, S. (ed), Compilation publication of CSCW'14 proceedings & CSCW'14 companion: February 15 - 19, 2014, Baltimore, Maryland, USA. New York, NY, ACM, pp. 229–232.

Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C. *et al.* (2012) Dynamic changes in motivation in collaborative citizen-science projects. In Poltrock, S., Simone, C., Grudin, J., Mark, G. and Riedl, J. (eds), the ACM 2012 conference, pp. 217-226.

Schweiz forscht (2021) What's Up, Switzerland? https://www.schweizforscht.ch/projekte/projektarchiv/what-s-up-switzerland.

Tiago, P., Gouveia, M. J., Capinha, C., Santos-Reis, M. and Pereira, H. M. (2017) The influence of motivational factors on the frequency of participation in citizen science activities. *Nature Conservation*, 18, 2017: 10.3897/natureconservation.18.13429.

Zampieri, M., Nakov, P. and Scherrer, Y. (2020) Natural language processing for similar languages, varieties, and dialects: A survey. *Natural Language Engineering*, 26, 2020: 10.1017/S1351324920000492.

7. Language Resource References

IamDiÖ. (2019). Wortgut, https://lex.dioe.at/