Developing Language Resources with Citizen Linguistics in Austria – A Case Study

Barbara Heinisch

Centre for Translation Studies, University of Vienna, Austria
Gymnasiumstraße 50, 1190 Vienna
barbara.heinisch@univie.ac.at

Abstract

Language resources are a major ingredient for the advancement of language technologies. Citizen linguistics can help to create language resources and annotate language resources, not only for the improvement of language technologies, such as machine translation but also for the advancement of linguistic research. The (language) resources covered in this article are a corpus related to the Question of the Month project strand, which was initially aimed at co-creation in citizen linguistics and a partially annotated database of pictures of written text in different languages found in the public sphere. The number of participants in these project strands differed significantly. Especially those activities that were related to data collection (and analysis) had a significantly higher number of contributions per participant. This especially held true for the activities with (prize) incentives. Nevertheless, the activities of the Question of the Month could reach a higher number of participants, even after the co-creation approach was no longer followed. In addition, the Question of the Month brought research gaps and new knowledge to light and challenged existing paradigms and practices. These are especially important for the advancement of scholarly research. Citizen linguistics can help gather and analyze linguistic data, including language resources, in a short period of time. Thus, it may help increase the access to and availability of language resources.

Keywords: Language varieties, citizen linguistics, language resource development

1. Introduction

The history of citizen linguistics in Austria looks back on a long tradition. Since citizen linguistics takes different forms, we may differentiate between citizens contributing to linguistic research that is coordinated and supervised by scholars, on the one hand, and so-called amateur linguists, on the other. Examples of activities by the latter are dictionaries compiled by people who are not trained lexicographers. This is because linguistics lends itself to the contribution by citizens since everybody uses language. This contribution goes beyond being a scholar's subject of investigation as speakers of a language (variety). It is rather about finding new research topics, data collection, data analysis or interpretation done by citizens according to scholarly principles.

1.1 History of Citizen Linguistics in Austria

Citizen linguistics in Austria dates back to the Habsburg Monarchy in the 19th century when it had a strong focus on the collection of linguistic data, especially of dialects. Two examples of these research initiatives in which citizens played an important role in collecting data from the actual speakers of dialects are the Dictionary of Bavarian Dialects in Austria (*Wörterbuch der bairischen Mundarten in Österreich*, WBÖ) and the Wenker Atlas.

In both cases, so-called amateur explorers were asked to empirically collect data of the local dialects. While the WBÖ was launched by two chancelleries in today's Germany and Austria, the Wenker-Atlas was initiated by Georg Wenker, who was a librarian in today's Germany.

1.1.1 Wörterbuch der bairischen Mundarten in Österreich (WBÖ)

The WBÖ was initiated with the aim to chart the Bavarian dialect region (*gesamtbairischen Dialektraum*) in a dialect dictionary. Since this endeavour was aimed at a comprehensive and systematic study of this dialect region, the scholars required help from volunteer data collectors who were recruited through newspaper announcements. The recruited explorers received written instructions for surveying the local population speaking the typical local dialect and collecting lexical data. Since then and over centuries, these data had been fed into the WBÖ dictionary (Stöckle, 2019; ÖAW-ACDH; WBÖ, 2020).

1.1.2 Wenker Atlas

The Wenker Atlas was aimed at finding the boundaries of dialects in the German Reich and at compiling the data in the *Sprachatlas des Deutschen Reichs* language atlas. To achieve the highest possible density of data collection points, local teachers served as explorers. They were tasked with the translation of the Wenker sentences that were written in standard German language into the local dialect. These data were then fed into the language atlas (Herrgen, 2010; DiWA, 2019).

In both cases, volunteers served as citizen linguists who collected data for linguistic research.

In the following section, the peculiarities of the Austrian variety of the German language are addressed to understand the background of the citizen linguistics project presented in this paper.

2. The Austrian Variety of the German Language

German is the official language in Austria, and it is a pluricentric language, "i.e. a language with several interacting centers, each providing a national variety with at least some of its own (codified) norms" (Clyne, 1995: 20). As a pluricentric language German has three standard varieties (Schmidlin, 2011), i.e. German, Austrian and Swiss. However, studies in the field of language geography

have shown that the German standards do not follow national borders but rather dialect boundaries (Elspaß et al, 2017). Therefore, the German language is rather a pluriareal (and not a pluricentric) language, making the collection and proper documentation of language resources for the Austrian variety more challenging.

The Austrian variety of the German language differs from the other varieties of German in several aspects (Wiesinger, 1988; Scheuringer, 2001), including lexical differences, pronunciation, the grammatical gender of nouns, the use of tenses or prepositions or the creation of diminutives or composita (Wiesinger, 1996). However, also within the Austrian standard variety differences between regions can be observed.

Moreover, language varieties in Austria, such as dialects are strongly related to a person's identity. Discussions about these varieties are, therefore, often ideological ones (Scheuringer, 1997; Cillia, 1995).

Within this framework, the citizen linguistics project "On everyone's mind and lips – German in Austria" was launched.

3. The Citizen Linguistics Project "On everyone's mind and lips – German in Austria"

The project "On everyone's mind and lips – German in Austria" (abbreviated as IamDiÖ in German) addresses the use and perception of the German language in Austria as well as the attitude of people towards it.

IamDiÖ consists of three project strands, each of which adopts another approach to citizen science. The first strand is entitled Question of the Month. It is aimed at co-creation which means that citizens can raise, and answer research questions related to the topic of German language in Austria. In defining the topic and question, selecting and applying methods to collect and/or analyze data and in interpreting the results, citizens should be supported by scholars, i.e. experts in the field of linguistics.

The second project strand addresses linguistic landscapes, which are defined as "the visibility and salience of languages on public and commercial signs in a given territory or region" (Landry and Bourhis, 1997, 23). Linguistic landscapes thus comprise street names, shop signs, billboard advertisements and stickers on lampposts, among others. A linguistic landscape serves different functions and may help to mark the relative status of linguistic communities in a certain region, among others (Landry and Bourhis, 1997). In order to be able to analyze a linguistic landscape, data in the form of pictures of written information in the public sphere, e.g. pictures of posters, shop signs or stickers on bicycle racks are needed. The third strand of the project is a meme contest, in which citizens generate data in the form of memes. Citizens are asked to combine text written in a dialect with pictures that can be associated with Austria. Since the creation of memes and their distribution via social media is rather an experiment than citizen science, this strand would not be regarded as citizen science, or rather citizen humanities, per se (Eitzel et al., 2017; Heigl et al., 2019).

In the following sections, the two citizen science strands are elaborated in more detail.

3.1 The Question of the Month

Co-creation is defined as public participation in scholarly research that sees citizens as co-researchers who are involved in any step and decision throughout the research process (Bonney et al., 2009). IamDiÖ intended to apply co-creation in the project's Question of the Month strand. This strand can be considered as a proof of concept for the idea of applying co-creation in citizen linguistics.

3.1.1 Co-creation in Citizen Linguistics

The idea behind the Question of the Month is that volunteers are involved and have a say in the entire research process. They are considered co-researchers. As the name of this project strand already suggests, it addresses research questions. These should be raised and, ideally, also be answered by citizens themselves. Researchers (only) support the volunteers in finding an answer to their questions, e.g. by helping select a method, suggest relevant literature or interpret the results. A Question of the Month should cover language use, language perception or language attitude with a focus on the German language in Austria, including all its varieties. Citizens can submit their questions via the IamDiÖ website. However, the number of questions collected during science communication events, such as the Long Night of Research in Austria or the Austrian Science Fund's Science and Society Festival, was tremendously higher, amounting to about 500 questions that were raised by citizens. These included question such as: "Do dialects in Austria disappear?", "Why do I have to face discrimination because I am from Germany and speak German German?" or "Does communication in social networks have a negative influence on 'good' German?". The volunteers who raised the questions were also asked if they would be willing to find an answer to their question. However, almost all of them refused to do research on their own, even if researchers offered their support. Therefore, the initial attempt of co-created research was foiled already in an early stage of the research process. This is also the reason why the co-creation approach could no longer be adopted in the project. Subsequently, the idea of the Question of the Month had to be re-considered as well.

3.1.2 From Co-creation to Science Communication

Instead of asking citizens to answer the research questions, the scholars in the project were required to respond to the questions. After all these questions had been collected from citizens, they were clustered according to topic. Every month, two questions per theme are selected by the project team. Here, the initial idea that two questions are selected, and in social networks citizens vote for the question that should be answered this month could still be put into action. After the users have voted for their favorite question, the question getting most of the votes is answered by the researchers. The scholars give an answer to the research question in a blog entry that follows a uniform structure. This structure reflects the research process and related steps, i.e. finding a topic, defining a research question,

doing a literature review, selecting a method, applying the method, analyzing data, writing about the results, interpreting the results and drawing conclusions. In this case, the conclusions are not only related to the research itself but also to the person and the personal development of the academic researcher (or the citizen humanist). This uniform structure that was oriented towards the research process should help readers gain an insight into the steps in the research process and increase academic literacy. As a final step, the scholar's (or citizen humanist's) answer is published as a blog entry on the IamDiÖ website and circulated via social networks. Interestingly, the questions raised by the citizens also helped to reveal research gaps.

Although, the citizens showed interest in the topic and raised a lot of questions in the initial project phase, this interest could not be sustained in the subsequent stages of the research process, thus, shifting the focus from co-creation to science communication in the other project phases.

3.2 Linguistic Landscaping

The second strand of the project can be regarded as collaborative approach to citizen science (Bonney *et al.*, 2009). This IamDiÖ strand is aimed at studying the linguistic landscape in Austria. Participants are asked to collect and analyze data in the form of pictures of written text in the public space, e.g. street names, posters or graffiti containing text. Citizens gather and analyze these pictures with the Lingscape app (Purschke, 2017; Seltmann and Heinisch, 2018).

3.2.1 Linguistic Treasure Hunts

To make linguistic landscape research more appealing to the participants, linguistic treasure hunts are organized in different cities in Austria. Linguistic treasure hunts as a method combine linguistic landscaping done by citizens with gamification. These are treasure hunts modified to the needs of citizen linguistics (with a focus on linguistic landscaping). Similar to treasure hunts in which a group of persons follows clues to get to a certain location, linguistic treasure hunts also have clues that are placed in an urban space and that participants have to solve to get to the next clue to finally win a prize. Since the groups move in the public space when they get from clue to clue, they also walk past written text. This text is interesting for linguistic landscape research, especially for research on language variation in writing. Therefore, with linguistic treasure hunts, scholars can pursue the objective of gathering data on and analyzing (written) language variation in the public sphere. In addition to the tasks completed in a traditional treasure hunt, the groups are tasked with taking, uploading and tagging photographs of written texts in the public sphere. The tagging task plays a crucial role since participants have to add annotations to the pictures, including geographical location, language(s) in which the text is written, language varieties, e.g. dialects, or function, medium and context. In linguistic treasure hunts, data quantity, i.e. the number of pictures uploaded and data quality, i.e. the annotation, have to be balanced: The groups do not only receive points for the number of uploaded

photographs but also for the tags (according to a point system). Finally, a prize is given to the group who followed all the clues, uploaded the most pictures and annotated them in accordance with predefined criteria (Heinisch, in print b).

3.2.2 Recruitment through Citizen Science Award

This project strand could recruit some participants through the Austrian Citizen Science Award, which is an event that helps citizen science projects recruit participants, i.e. school classes and individuals. Within a specified period of time, these classes and individuals can contribute to a range of citizen science projects. These contributions can be data collection, data analysis, etc. The most successful classes and persons receive prizes from each citizen science project in a festive ceremony.

For linguistic landscaping, the instructions for the participants were to take pictures of written text in the public space and upload, geolocate and tag them with the *Lingscape* app. The individuals with the highest number of pictures uploaded (and tagged) win the prize, whereas the class with the highest amount of uploaded (and tagged) pictures and who, additionally, submitted a research report receives the prize.

4. Language Resources

The language resources created by these two project strands address the diversity of the Austrian variety of the German language and the diverse use of language(s) in Austria.

First, the language resource comprising the Questions of the Month (IamDiÖ, 2019) is a corpus of questions and answers addressing the Austrian variety of the German language. These questions and answers range from the use of language(s) and their varieties in Austria, language change, perception of and attitudes towards language(s) and their varieties. While this monolingual corpus has a clear thematic focus on the Austrian variety of the German language, the corpus itself is in both Austrian and German standard varieties since the academics (and citizen humanists) writing the answers have diverse language backgrounds. Although this corpus is not annotated, it has a clear structure. As mentioned before, the corpus consists of questions and answers according to a predefined structure derived from the steps in the scholarly research process. This monolingual written corpus in German is available under a Creative Commons licence. It is newly created and constantly added to. This language resource lends itself to information retrieval and extraction, knowledge discovery or representation or machine learning.

Second, the data collected through the linguistic treasure hunts may not be regarded as language resource *sensu stricto*, since the pictures containing text are only available as pictures (IamDiÖ & Lingscape, 2019). Optical character recognition has not been used so far, but the pictures are annotated according to an annotation scheme, which was developed by the IamDiÖ team for the linguistic treasure hunts (Heinisch, in print b). The pictures and annotations

made during the linguistic treasure hunts were integrated into the *Lingscape* database, which is a (partially) annotated database of photographs of text written in different languages found in the public sphere. This database is, therefore, a compilation of pictures and annotations from different projects aimed at the analysis of linguistic landscapes in different countries. To make this resource available for further use, e.g. natural language processing, it would need further preparatory work.

5. Comparison of Collaborative and Co-created Project Strands

A comparison of the two project strands focusing on citizen linguistics should reveal the success of each. However, a comparison proved challenging not only because each citizen science project defines success differently (Freitag and Pfeffer, 2013), but also due to the different approaches and topics of these strands. The criteria used for the comparative analysis were the number of participants, the number of contributions (per participant) and perceived advancement in scholarship (Heinisch, in print a). It must be noted that this study was not planned in advance. It was only implemented after the first phase of the project ended. This means that no rigid data collection principles had been defined beforehand, but all the available data (including estimations) were aggregated only afterwards to answer the question of which project strand was more successful.

5.1 Criteria

Despite the ongoing debate on success in the citizen science literature and criteria defined (Cox *et al.*, 2015; Freitag and Pfeffer, 2013), the available data made it necessary to specify own criteria, namely the number of participants, the number of contributions per participant and perceived advancement in scholarship (Table 1). The number of participants had to be partly estimated since no rigid counting of science festival visitors was applied. The (average) contributions per participant are based on the overall number of contributions and the (estimated) number of participants. Contributions to the Question of the Month project strand are the (average) number of research questions raised per participant, whereas contributions to the linguistic treasure hunts are the (average) number of pictures uploaded to the app. The perceived advancement to scholarship is based on the author's personal perception of the contribution of each of the activities to scholarly knowledge or academia in general. Finally, Table 1 also contains information on the degree of voluntariness, which will be elaborated later (Heinisch, in print a).

5.2 Comparison

The comparative analysis (Heinisch, in print a) demonstrated that the project strand aimed at co-creation attracted more participants overall (but only in the initial research phase in which the task was to find a research question) (Table 1). This is in contrast to the number of contributions per participant that were significantly higher for the linguistic treasure hunts. These differences in numbers may be attributed to various factors. The most obvious one is that the topic of German in Austria was appealing to a high number of people and the data, i.e. the research questions for the Question of the Month were collected from visitors of science communication festivals based on personal dialogue. This allowed for the collection of about 500 questions in total. The comparison between the Question of the Month and the linguistic treasure hunt demonstrated that the task of crowdsourcing, i.e. soliciting contributions from the crowd, i.e. a large group of unfamiliar individuals (Bowser and Shanley, 2013), yielded the better results regarding data quantity (Heinisch, in print a).

Another category in which the project strands were compared was the degree of voluntariness, which can be related to a person's motivation for participating in a certain citizen science activity. The practice of involving school classes or university students in citizen science, raises the issue of voluntary participation, since the citizen science tasks are often mandatory parts in a school subject or university course.

According to the Oxford English Dictionary (2020), voluntariness is "[t]he state or condition of being voluntary, free, or unconstrained; absolute freedom or liberty in respect of choice, determination, or action". In addition to openness and collaboration, voluntariness is one of the basic ideas in citizen science (Fresa and Justrell, 2015). Therefore, the study (Heinisch, in print a) differentiated between three degrees of voluntariness, i.e. voluntary (the participants freely decided to participate in the task at hand, e.g. based on their interest in the topic), semi-voluntary (the participants were given an incentive to participate, but the decision to take part in the activity was taken freely) and non-voluntary (which includes some type of compulsion). This categorisation shows a strong link to the debate on intrinsic and extrinsic motivation. It is assumed that especially non-voluntary participation may negatively affect motivation, data quality and data quantity. However, these needs to be further investigated.

When comparing the Question of the Month and the linguistic treasure hunt from the point of view of voluntariness, the Question of the Month boosts a higher degree of voluntariness, since the majority of the questions were raised out of curiosity. As the questions were primarily collected during science communication events, the citizens' contributions can be considered voluntary ones since only people who are interested in the topic enter a project's festival booth. Nevertheless, also the Question of the Month strand had some semi-voluntary contributions, since university students were encouraged to deliver questions and/or answers. Here, for some university students the submission of research questions was a mandatory part of a course. In other university courses it was no compulsory assignment but a semi-voluntary one, since students could get bonus points for a course. In general, only one participant (from the bonus point group) was willing to answer her own research question.

For the linguistic treasure hunts, which were organized several times in Austrian cities throughout the project, semi-voluntary participation prevailed. This is due to the

fact that the majority of the participants were university students receiving bonus points.

While we can assume that participation of individuals in the Citizen Science Award is semi-voluntary, and either driven by intrinsic motivation or the prize incentive, the participation of the school classes can be regarded as semi-voluntary (the teachers may participate out of interest in the topic and/or to win a prize for the class; but their class must participate since the citizen science activities are part of the relevant subject at school).

In general, the number of pictures uploaded was higher if there was an incentive, either bonus points for university students or a prize. This increase in data quantity due to the prize incentive especially held true for the individuals who participated in the Citizen Science Award competition.

The contributions to the advancement in scholarship differ significantly between the two project strands. While the linguistic treasure hunts could primarily increase the amount of (partially) annotated data for linguistic landscaping research, the Question of the Month strand revealed knowledge and research gaps, helped raise new questions, challenged established approaches in academia and questioned paradigms (in scholarly research). Since one participant found an answer to her research question without the help of scholars, but according to the principles of academic research, also independent research could be observed.

6. Discussion

There is a growing body of literature that recognizes motivation in citizen science (Moczek, 2019; Oded Nov, Ofer Arazy, David Anderson, 2011; Raddick *et al.*, 2010), but far too little attention has been paid to the voluntariness of participation. Studies of gamification in citizen science show the importance of data quality and motivation (Tinati *et al.*, 2017; Curtis, 2015; Prestopnik and Crowston, 2011). Gamification was also an inherent part of the linguistic treasure hunts. Gamification, which is accompanied by competition, helped to strengthen the motivation of treasure hunt participants and increased the amount of data gathered, but it also may impede data quality, especially the quality of the annotations (Heinisch, in print b). Finding the right balance between data quantity and data quality is also a major area of interest in citizen science (Bordogna *et al.;* Crall *et al.;* Ellwood *et al.*, 2016; Hunter *et al.*, 2013; Kelling *et al.;* Kosmala *et al.*, 2016; Prats López, 2017). Means of quality control and evaluation could also help to increase the quality of the data gathered during linguistic treasure hunts.

7. Conclusion

Language resources are a major ingredient for the advancement of language technologies. Citizen linguistics can help to create language resources and annotate language resources. This is important not only for the improvement of language technologies, such as machine translation but also for the advancement of linguistic research.

Exemplified by the citizen linguistics project "On everyone's mind and lips – German in Austria", two approaches to citizen linguistics were compared, i.e. an attempt to implement co-creation in the citizen humanities (the Question of the Month) on the one hand, and a collaborative approach to linguistic landscaping (including linguistic treasure hunts), on the other. The (language) resources created by these two approaches are a corpus related to the Question of the Month project strand and a partially annotated database of pictures of written text in different languages and language varieties found in the public sphere.

The number of participants in these two project strands differed significantly. Especially those activities that were related to data collection (and analysis) had a significantly higher number of contributions per participant. This especially held true for the activities with (prize) incentives. Nevertheless, the activities of the Question of the Month that aimed at co-creation could reach a higher number of participants, even after the co-creation approach was no longer followed. In addition, especially the Question of the Month brought research gaps and new knowledge to light and challenged existing paradigms and practices.

Citizen linguistics can help gather and analyze linguistic data, including language resources, in a short period of time. Thus, it may help increase the access to and availability of language resources, including language resources particular to a certain language variety, e.g. language resources in standard varieties or dialects. Therefore, citizen linguistics can play a crucial role in the advancement of language technologies and scholarly research.

Project strand	Communication	Number of participants	Number of contributions per participant	Contribution to advancement in scholarship	Voluntariness/motivation
Question of the Month (QM)	QM Festivals	350 (estimation)	1-5 (estimation)	New research topics Challenging established approaches/paradigms	Voluntary/interest
	QM university courses	20 (two universities)	1	Partly independent research into their individual questions	Incentive: part of the course or bonus points for the course
	QM web form and e-mail	4	4	New research topics Challenging established approaches/paradigms	Voluntary/interest
Linguistic landscaping (LL)	LL treasure hunts	20 (two cities)	16 (on average) (with prize: 29; without prize 7)	Data collection and initial analysis	Voluntary (4 persons) Bonus point for course (16 persons) Incentive: prize vs no prize
	LL Austrian Citizen Science Award	4 registered individuals 7 registered school classes	83 (individual) 38 (school)	Data collection and initial analysis Partly: new research topics	Incentive: prize

Table 1 : Comparison of the two project strands Question of the Month and linguistic landscaping (in July 2019)

9. Bibliographical References

Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J. et al. (2009) Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. http://files.eric.ed.gov/fulltext/ED519688.pdf (last accessed February 11, 2016).

Bordogna, G., Carrara, P., Criscuolo, L., Pepe, M. and Rampini, A. A linguistic decision making approach to assess the quality of volunteer geographic information for citizen science. Information Sciences, 10 February 2014, Vol.258, pp.312-327, 312.

Bowser, A. and Shanley, L. A. (2013) New Visions in Citizen Science. https://www.wilsoncenter.org/sites/default/files/NewVisionsInCitizenScience.pdf (last accessed August 28, 2017).

Cillia, R. de (1995) Deutsche Sprache und österreichische Identität. Medienimpulse, 4, 4–13.

Cox, J., Oh, E. Y., Simmons, B., Lintott, C., Masters, K., Greenhill, A. et al. (2015) Defining and Measuring Success in Online Citizen Science. A Case Study of Zooniverse Projects. Computing in Science & Engineering, 17, 2015: 10.1109/MCSE.2015.65.

Crall, A. W., Newman, G. J., Stohlgren, T. J., Holfelder, K. A., Graham, J. and Waller, D. M. Assessing citizen science data quality: an invasive species case study. Conservation Letters, 2011, Vol.4(6), pp.433-442, 433.

Curtis, V. (2015) Motivation to Participate in an Online Citizen Science Game. Science Communication, 37, 2015: 10.1177/1075547015609322.

DiWA (2019) Die Rolle des Wenker-Atlasses in der Geschichte der Dialektologie. http://www.diwa.info/Geschichte/RolleDesWenkeratlasses.aspx.

Eitzel, M. V., Cappadonna, J. L., Santos-Lang, C., Duerr, R. E., Virapongse, A., West, S. E. et al. (2017) Citizen Science Terminology Matters. Exploring Key Terms. Citizen Science: Theory and Practice, 2, 2017: 10.5334/cstp.96.

Ellwood, E., Henry Bart, JR, Doosey, M., Jue, D., Mann, J., Nelson, G. et al. (2016) Mapping Life – Quality Assessment of Novice vs. Expert Georeferencers. Citizen Science: Theory and Practice, 1, 2016: 10.5334/cstp.30.

Freitag, A. and Pfeffer, M. J. (2013) Process, not product: investigating recommendations for improving citizen science « success ». PloS one, 8, 2013: 10.1371/journal.pone.0064079.

Fresa, A. and Justrell, B. (2015) Roadmap for Citizen Science. https://www.civic-epistemologies.eu/wp-content/uploads/2014/08/CE_Roadmap-Handbook.pdf.

Heigl, F., Kieslinger, B., Paul, K. T., Uhlik, J. and Dörler, D. (2019) Opinion: Toward an international definition of citizen science. Proceedings of the National Academy of Sciences of the United States of America, 116, 2019: 10.1073/pnas.1903393116.

Heinisch, B. (in print a) Comparison of co-created and collaborative approaches to citizen science adopted by the citizen linguistics project 'On everyone's mind and lips – German in Austria'. In, Proceedings of the 5th Austrian Citizen Science Conference 2019, 26-28, June, 2019, Obergurgl, Austria.

Heinisch, B. (in print b) Hunting for signs in the public space – the method of linguistic treasure hunts as a form of citizen science. In, Proceedings of the 5th Austrian Citizen Science Conference 2019, 26-28, June, 2019, Obergurgl, Austria.

Herrgen, J. (2010) The digital wenker atlas (www.diwa.info): an online research tool for modern dialectology. Dialectología, 95.

Hunter, J., Alabri, A. and Ingen, C. (2013) Assessing the quality and trustworthiness of citizen science data. Concurrency and Computation: Practice and Experience, 25, 454–66.

Kelling, S., Fink, D., Sorte, F., Johnston, A., Bruns, N. and Hochachka, W. Taking a 'Big Data' approach to data quality in a citizen science project. Ambio, 2015, Vol.44(4), pp.601-611, 601.

Kosmala, M., Wiggins, A., Swanson, A. and Simmons, B. (2016) Assessing data quality in citizen science. Frontiers in Ecology and the Environment, 14, 2016: 10.1002/fee.1436.

Landry, R. and Bourhis, R. Y. (1997) Linguistic Landscape and Ethnolinguistic Vitality. Journal of Language and Social Psychology, 16, 1997: 10.1177/0261927X970161002.

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.

ÖAW-ACDH Wörterbuch der bairischen Mundarten in Österreich (WBÖ). https://vawadioe.acdh.oeaw.ac.at/projekte/wboe/wboe-startseite/.

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

Oxford English Dictionary (OED) (2020)Voluntariness, Oxford University Press.

Prats López, M. (2017) Managing Citizen Science in the Humanities: The challenge of ensuring quality, Vrije Universiteit. http://hdl.handle.net/1871/55271.

Prestopnik, N. R. and Crowston, K. (2011) Gaming for (Citizen) Science: Exploring Motivation and Data Quality in the Context of Crowdsourced Science through the Design and Evaluation of a Social-Computational System. In, 2011 IEEE Seventh International Conference on e-Science Workshops. IEEE, pp. 28–33.

Purschke, C. (2017) Crowdsourcing the linguistic landscape of a multilingual country. Introducing Lingscape in Luxembourg, 2017: 10.13092/lo.85.4086.

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.

Scheuringer, H. (1997) Sprachvarietäten in Österreich. In Stickel, G. (ed), Varietäten des Deutschen: Regional- und Umgangssprachen. Berlin, New York, de Gruyter, pp. 332–345.

Seltmann, M. E.-H. and Heinisch, B. (2018) How to speak German in Austria. Collaboration between two linguistic citizen science projects – "On everyone's

mind and lips – German in Austria" and "Lingscape" found each other, FRONTIERS MEDIA SA. https://klf.univie.ac.at/fileadmin/user_upload/p_klf/Proceedings_OECSK2018.pdf, 82–85.

Stöckle, P. (2019) Wie ein Dialektwörterbuch entsteht. https://dioe.at/details/artikel/1963/.

Tinati, R., Luczak-Roesch, M., Simperl, E. and Hall, W. (2017) An investigation of player motivations in Eyewire, a gamified citizen science project. Computers in Human Behavior, 73, 2017: 10.1016/j.chb.2016.12.074.

WBÖ (2020) Materialbasis. https://vawadioe.acdh.oeaw.ac.at/projekte/wboe/materialbasis/.

10. Language Resource References

IamDiÖ. (2019). Frage des Monats, https://iam.dioe.at/frage-des-monats/beantwortete-fragen/

IamDiÖ & Lingscape (2019). Lingscape https://lingscape.carto.com/builder/781d0814-ef0d-11e6-ad6f-0e3ff518bd15/