

# Dia-Lingle: A Gamified Interface for Dialectal Data Collection

Jiugeng Sun<sup>1</sup>, Rita Sevastjanova<sup>1\*</sup>, Sina Ahmadi<sup>2\*</sup>,  
Rico Sennrich<sup>2</sup>, Mennatallah El-Assady<sup>1</sup>

<sup>1</sup>Department of Computer Science, ETH Zürich, Switzerland

<sup>2</sup>Department of Computational Linguistics, University of Zürich, Switzerland

Correspondence: [jiusun@ethz.ch](mailto:jiusun@ethz.ch)

## Abstract

Dialects suffer from the scarcity of computational textual resources as they exist predominantly in spoken rather than written form and exhibit remarkable geographical diversity. Collecting dialect data and subsequently integrating it into current language technologies present significant obstacles. Gamification has been proven to facilitate remote data collection processes with great ease and on a substantially wider scale. This paper introduces Dia-Lingle, a gamified interface aimed to improve and facilitate dialectal data collection tasks such as corpus expansion and dialect labelling. The platform features two key components: the first challenges users to rewrite sentences in their dialects, identifies them through a classifier and solicits feedback, and the other one asks users to match sentences to their geographical locations. Dia-Lingle combines active learning with gamified difficulty levels, strategically encouraging prolonged user engagement while efficiently enriching the dialect corpus. Usability evaluation shows that our interface demonstrates high levels of user satisfaction. We provide the link to Dia-Lingle: <https://dia-lingle.ivia.ch/>, and demo video: <https://youtu.be/0QyJsB8ym64>.

## 1 Introduction

Dialects present unique challenges for computational linguistics due to their scarcity of textual resources, with limited datasets and digital documentation tools dedicated to their study and analysis (Joshi et al., 2024). This digital resource gap, which affects low-resourced languages broadly, reflects systematic issues like cultural marginalisation. Consequently, language communities are often unable to fully benefit from advancements in language technology, raising concerns about the potential erasure of linguistic and cultural diversity in the AI models (Grützner-Zahn et al., 2024).

In the current AI landscape, where data plays a central role—particularly in the advancement of large language models (LLMs)—ensuring fair representation of dialects remains a pressing challenge. Paradoxically, the pragmatic implementation of language standards often forces dialects into standardised written forms for efficiency, which fundamentally undermines authentic dialectal representation and complicates data collection efforts (Auer, 2005). Previous approaches in natural language processing (NLP) have attempted to address this challenge by leveraging syntactic atlases, structured questionnaires, and direct annotation by native speakers, though these annotations remain limited in scope, as discussed by Alam et al. (2024).

In this vein, we design and develop Dia-Lingle, an interactive dashboard that aims to collect dialectal data through two gamified components, as showcased in Figure 1. In the first component, dubbed ‘Quiz’, users rewrite standardised sentences, i.e. sentences in the standard variety of the language, in their dialects. A dialect identification classifier predicts the geographical origin of these rewritten sentences followed by a visualisation of these dialect regions to users, soliciting their feedback both to refine classification accuracy and to build a more comprehensive dialectal dataset. In addition, we integrate active learning (AL) techniques for the strategic recommendation of sentences, selecting suitable ones based on model uncertainty measurements and explicit user feedback. This AL approach is embedded within a difficulty level setting, where challenges escalate based on user proficiency and model performance, thereby enhancing engagement and extending participation in the data collection cycle. The second component, dubbed ‘Match’, requires users to match example sentences to the geographical areas where they are spoken. A comprehensive usability study is conducted, revealing user satisfaction from interface design to overall concept.

\*Equal contribution.



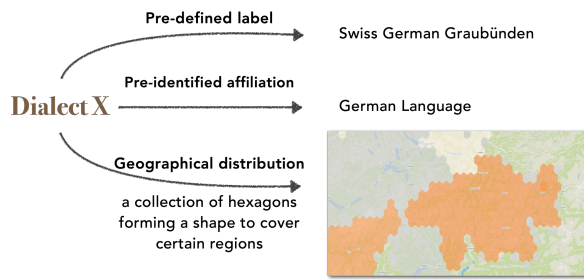


Figure 2: Illustration of dialect representation in Dia-Lingle using Dialect X as an example spoken primarily in the Graubünden region of Switzerland.

### 3.1 Users

Dia-Lingle is designed for three distinct user groups: (a) dialect community members with basic to proficient understanding of specific dialects or dialect families; (b) general linguists interested in analysing the classifier performance and providing sophisticated feedback; and (c) learners in educational contexts to engage with Match component to better understand dialectal variations. This multi-audience approach ensures the platform serves both data collection and model refinement purposes.

### 3.2 Data Structure

Dia-Lingle utilises two different types of datasets that serve distinct functions within the system. The first dataset comprises archival general information about each dialect, which enriches the user experience by providing contextual background and enhances the visualisation component of the interface. The second dataset contains a corpus of sentences written in various dialects, which provides the training data for the dialect identification classifier and forms the foundation of an expanding resource that will be aggregated for future downstream machine learning tasks like dialect-specific language modelling. Both datasets are continually refined through user feedback embedded in the interface, creating a dynamic and increasingly comprehensive resource for dialectal research.

#### 3.2.1 Dialect Representation

Our approach represents dialects from data science and geographical perspectives rather than discussing their political factors or sociolinguistic impact. Each dialect in our database is assigned three attributes: a pre-defined label, a pre-identified affiliation (i.e., to which macro-language family it belongs), and a geographical distribution. The geographical distribution is represented as a polygon

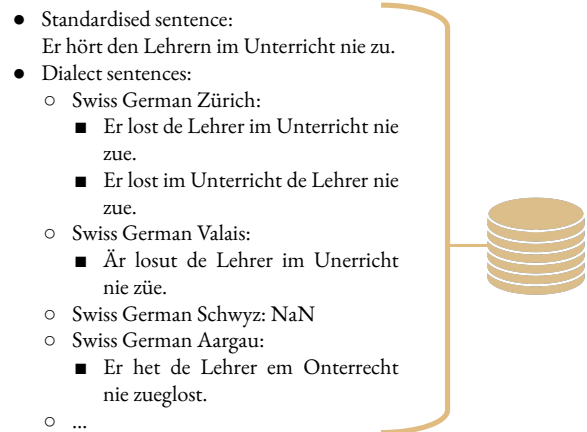


Figure 3: Illustration of a parallel sentence group in Swiss German. There is one standardised sentence and multiple dialect sentences that convey the exact same meaning.

which is formed by a collection of hexagons that covers the regions where the dialect is predominantly spoken. The hexagon sizes are tailored to correspond with the geographical coverage of each language on the world map, with smaller hexagons representing languages spoken in more limited regions, so that the collective hexagon arrangement forms a polygon shape reflecting each language’s dialectal distribution. Figure 2 illustrates this approach with an example of *Dialect X*, which is primarily spoken in the Graubünden region of Switzerland.

#### 3.2.2 Corpus

The corpus comprises a collection of parallel sentences, with each data entry containing a standardised language version alongside dialectal variations that convey the same meaning. All sentences are labelled using the same pre-defined labels established in the dialect representation dataset as discussed in Section 3.2.1. Importantly, a dialectal sentence may belong to more than one dialect category, reflecting the natural overlap in linguistic features across related dialect groups. We present an example on one of the possible dialect sentence groups in Swiss German to better illustrate the corpus structure in Figure 3.

Currently, our corpus encompasses five language families: Swiss German (2,760 parallel sentence groups with eight dialect variations), Kurdish (300 parallel sentence groups with four dialect variations), Korean (51,963 parallel sentence groups with one dialect variation), Japanese (500 parallel sentence groups with twenty-one dialect variations)

and Romansh (113 parallel sentence groups with five dialect variations). These datasets were established from existing resources including SwissDial (Dogan-Schönberger et al., 2021), CODET (Alam et al., 2024), Jejeuo Dataset (Park et al., 2020), and CPJD corpus (Takamichi and Saruwatari, 2018). For Romansh, we relied on a few stories from Storyweaver<sup>1</sup> and their translations by the *Romanische Kindergeschichten* application (all licensed CC BY 4.0).

### 3.3 Classification

For dialect classification, Dia-Lingle relies on Meta’s fasttext (Mikolov et al., 2018) which provides character-level word embeddings, sentence classification and most importantly, language identification for 157 languages. For the task of dialect classification, we train models for the selected varieties in our corpus. During training, we use automatic hyperparameter optimisation, constraining the model size to a maximum of 2 MB and limiting the autotune duration to 500 seconds. The model performance was evaluated using an 80/20 train-test split. The  $F_1$  scores for different dialects are as follows: 0.878 for Swiss German, 0.546 for Kurdish, 0.973 for Korean, 0.446 for Japanese and 0.842 for Romansh. Some language families still have relatively low  $F_1$  scores due to limited corpus resources. Given that the interface focuses on integrating human feedback for data collection, model performance is less prioritised at this moment.

### 3.4 Gamification on Progressive Levels

Game interface design encompasses a diverse array of design patterns, among which the Levels are a prominent exemplar (Deterding et al., 2011). Levels are a mechanism to provide users with progression within the system (Hallifax et al., 2023). The gamified strategies include offering rewards for completing levels (Toda et al., 2019) or increasing difficulty (Seaborn and Fels, 2015). Snyder and Hartig (2013) develop a voluntary online quiz system and reveal through post-questionnaires that 96% of participants attribute their participation to difficulty level settings. In Dia-Lingle, we implement a three-tiered difficulty level setting: Easy, Normal, and Hard. Users initially receive a standardised sentence from the Easy category when entering the rewriting page. Either they can modify the difficulty level themselves, or upon expressing

satisfaction with the classification result and dialect visualisation, Dia-Lingle increases the difficulty level to encourage further participation.

### 3.5 Uncertainty-based Sample Selection

During data aggregation process, selecting sentences for users to rewrite requires careful consideration. Our objectives are twofold: to collect valuable and preferably unseen dialectal data, whilst encouraging user engagement. In each interaction round, we define a difficulty score  $D(s)$  for a standardised sentence  $s$  as:

$$D(s) = \sum_{k \in K} H(k) \quad (1)$$

where  $k$  denotes the  $k$ -th dialect variation class and  $H(k)$  represents the classification entropy of the  $k$ -th class for a standardised sentence  $s$ . We define the parallel dialect sentence group of sentence  $s$  to be  $G_s$ , where  $G_{s,k} \subset G_s$  denotes the subset of dialect sentences with label  $k$ . Furthermore, we define  $K$  as the set of all possible dialect variations and  $C \subseteq K$  as the set of labels that actually appear in  $G_s$ . We define  $H(k)$  to be:

$$H(k) = \begin{cases} -\frac{1}{|G_{s,k}|} \sum_{n \in G_{s,k}} \sum_{i \in K} p_{n_i} \log(p_{n_i}) & k \in C \\ H_{\max}(k) = -\sum_{i \in K} \frac{1}{|K|} \log(\frac{1}{|K|}) & k \notin C \end{cases} \quad (2)$$

where  $p_{n_i}$  is the probability of the dialect sentence  $n$  belonging to the  $i$ -th class. This score increases when certain dialect variations are missing from the parallel group or when the model struggles to confidently classify the dialect sentences. Standardised sentences are categorised by difficulty score: Hard (top 20%), Normal (middle 60%), and Easy (bottom 20%). As users interact with the Dia-Lingle interface, we continuously expand the corpus by adding more dialect sentences to the corresponding group  $G_s$ , potentially expanding the label set  $K$  and updating the dialect classifiers.

## 4 Interface Design

Dia-Lingle’s interface, depicted in Figure 4, guides users through several stages: Entry, Choice, Quiz, Review and Match. These stages encompass two distinct and separate gamified components that users can engage with. In our commitment to engage with local dialect communities, we have prioritised multilingual support across the interface. Beyond the interactions occurring within the main user flow, the interface incorporates numerous additional interactive components to enhance usability,

<sup>1</sup><https://storyweaver.org.in>



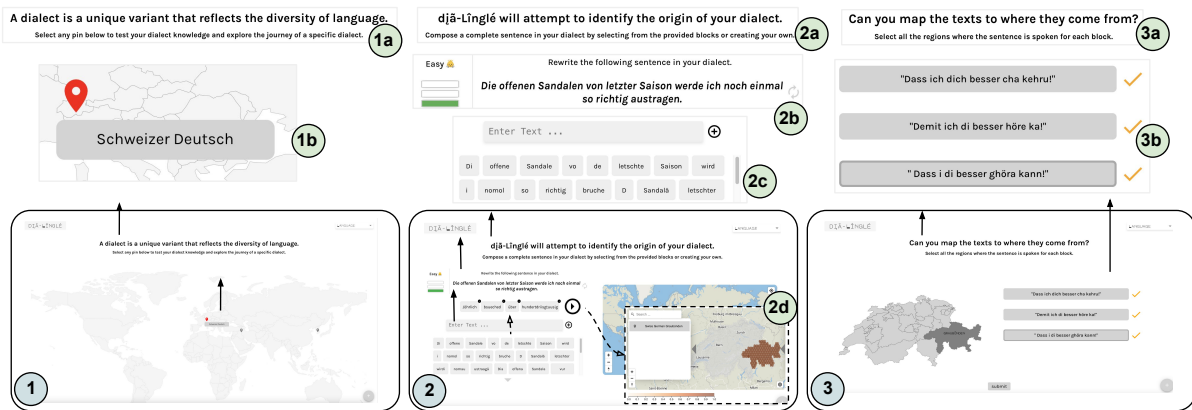


Figure 4: Simplified overview of Dia-Lingle interface design, detailed in Section 4. Major components are enlarged for visibility and labelled with circled numbers for reference. Appendix A provides additional images of other stages.

which include pop-up windows to prevent unintended actions or onboarding guidelines.

#### 4.1 Entry

Upon entering Dia-Lingle, users are greeted in a welcome page featuring a centred title as shown in Figure 4 (1a). Beneath is the navigation area, a world map displays several pins representing the current language families supported by the system. Users progress by selecting any of these pins as depicted in Figure 4 (1b).

#### 4.2 Choice

After clicking any pin on the world map, users enter the Choice page divided by a vertical dashed line. The two blocks positioned on either side represent distinct paths leading to two different gamified components as shown in Figure 5. This page requires users to specify whether they are familiar with the language they have selected. Their response determines which of the two major gamified components they will access.

#### 4.3 Quiz

Users proficient in the language access a page to rewrite sentences in their dialect. The interface



Figure 5: Illustration of the two gamified components (Quiz and Match) as they appear on the Choice page.

includes a title (Figure 4 (2a)), a rewriting dashboard with the standardised reference sentence and difficulty toggle (Figure 4 (2b)), and below these, a text input field with word suggestions from the dialect corpus (Figure 4 (2c)). Additionally, users can compose their dialectal sentence either by typing directly or by selecting suggested words. The sentence is assembled using interactive blocks that allow users to rearrange word order or remove elements as needed. The interface accommodates left-to-right and right-to-left writing systems. Once the rewriting is complete, users submit their sentence, triggering the classifier to identify the dialectal labels of the input. The classification results are then visualised on a map as shown in Figure 4 (2d).

#### 4.4 Review

The moment the map visualises the classification result, the rewriting interface on the left disappears, replaced by a question asking users whether they believe the identification is accurate. If users are satisfied with the prediction, the interface increases the difficulty level and redirects them back to the rewriting interface described in Section 4.3. Otherwise, the interface initiates a feedback collection process. It firstly displays a comprehensive list of all currently archived dialects that belong to the selected language. Users may then select the option that best matches their dialect, or choose *none of the above* to create their own entry. Importantly, regardless of which option users select, they can only modify the geographical regions associated with the dialect. They cannot alter either the dialect labels or the language affiliations detailed in Section 3.2.1. This restriction acknowledges the

sensitivity surrounding dialect and language naming conventions.

Following the users' selection, the interface activates a lasso tool enabling them to colour the areas on the map where they believe the dialect is spoken. Users can add or remove highlighted small hexagons to form a new polygon representation through either group selection or individual clicking. Upon submission of these geographical modifications, users are redirected back to the rewriting page without an increase in difficulty level.

The detailed illustrations of the Review session are included in [Appendix A](#).

#### 4.5 Match

Users who express unfamiliarity with the language in Section 4.2 are directed to an alternative component. In this gamified component, the interface presents three sentences sequentially as illustrated in [Figure 4 \(3b\)](#), asking users to highlight regions where they believe each sentence might be spoken. Once the reference answers are revealed, users have the chance to correct these mappings if they disagree with the suggested geographical distributions.

To maintain simplicity and intuitive gameplay, this matching exercise operates at the administrative division level (such as cantons or provinces) rather than using the more granular hexagonal mapping system. This design choice ensures Match component remains accessible to people of all backgrounds, regardless of their linguistic expertise.

### 5 Evaluation

We conducted a user study on the interface design discussed in Section 4. The comprehensive experimentation on the interface was carried out through surveys containing basic demographic data questions and questions following the System Usability Scale (SUS) format ([Brooke, 1995](#)). We also conducted usability studies to gain a better understanding of the participants' comprehension of the entire workflow and any potential obstacles they might encounter whilst interacting with Dia-Lingle. The contents of the survey and the detailed process of the usability test are illustrated in [Appendix A](#).

**Quantitative Evaluation** In total, 18 participants interacted with the interface and completed the survey. Among the participants, 13 people reported using their native languages to explore the interface. We calculate the overall average SUS score

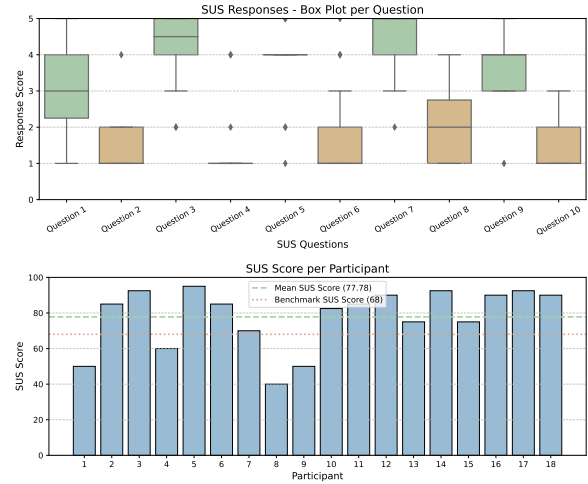


Figure 6: Analysis of SUS results for Dia-Lingle. Top: Box-plot of SUS scores by individual questions, with green boxes indicating higher values are better and brown boxes indicating lower values are better. Bottom: Bar chart showing individual user SUS scores. SUS questions are provided in [Appendix A, Table 1](#).

to be 77.78, as defined by [Brooke \(1995\)](#). Detailed results of SUS scores by question and individual user can be found in [Figure 6](#).

In the SUS survey, most questions received highly satisfactory results. However, the first question *I think I would use this interface frequently* produced mixed results. Although most participants expressed appreciation for the difficulty levels that motivated them play for more rounds, some preferred elements like point collection or competition for a long-lasting engagement. This supports previous research indicating that game element preferences are subjective ([Tondello et al., 2016](#)), highlighting the need for more personalised solutions. Based on the demographic data gathered, non-native speakers highly recognised the educational value. Among native speakers, Swiss German speakers reported higher satisfaction compared to Korean participants, attributable to the more comprehensive pre-established Swiss German dataset, which fostered greater trust and gave users more confidence in utilising the interface.

**Qualitative Evaluation** We conducted usability studies with four participants in total: three Swiss German native speakers and one Korean native speaker. The study instructions and a detailed analysis of identified strengths and problems during the evaluations are provided in [Appendix A](#). Overall participants exhibited a notably positive reception to the interface's core concept, with par-

ticular emphasis on the gamification that substantially increased the platform’s attractiveness and user engagement to them.

One key finding is the critical importance of correctly classifying the first rewritten sentence. Misclassification at this stage raised doubts among users, significantly reducing their confidence in proceeding. Another major finding during the studies is that while the difficulty scores described in Section 3.4 are based on the ML model’s uncertainty and missing data entries, participants reported that higher difficulty levels actually corresponded to increased complexity in rewriting standardised sentences to them. This was due to either the sentences themselves being inherently more challenging or the suggestion text blocks becoming fewer. Although it remains unclear whether the backend AL algorithm or the frontend difficulty level visualisation played a more significant role in user perception, the overall gamification approach successfully increased participant engagement. These findings offer valuable insights for the future development of Dia-Lingle.

## 6 Conclusion

Dia-Lingle introduces a novel gamified approach to dialectal data collection, addressing the critical challenge of resource scarcity for dialects in computational linguistics. By integrating active learning with gamification elements, our platform creates an engaging environment that encourages prolonged user participation while systematically enriching dialectal corpora. The two-component design—Quiz for dialect rewriting and identification, and Match for geographical mapping—provides complementary pathways for data collection tailored to different user knowledge levels. Our usability evaluation demonstrates high levels of user satisfaction with the interface design and overall concept. The integration of difficulty progression, geographical visualisation, and interactive feedback mechanisms has proven effective in sustaining user engagement. Key insights from our evaluation highlight the importance of accurate initial classification in establishing user trust and the value of strategically increasing challenge levels to maintain participation.

Dia-Lingle represents a significant step toward more inclusive language technology by creating pathways for communities to contribute directly to the resources that will power future NLP applications. This, consequently, also contributes to the

broader goal of preserving linguistic diversity.

## Limitations and Future Work

Currently we only support five different languages on the dashboard. In future research, we aim to broaden the scope of language coverage in dialect identification. As migration and language contact influence speakers, sentences may exhibit characteristics of multiple dialects within a single language or even features from multiple languages. Beyond probability distributions, more sophisticated visualisation methods are needed to effectively represent dialectal mixtures. Additionally, developing machine learning models capable of encoding and detecting such patterns remains an open research challenge. Furthermore, dialects are an integral part of language diversity, and they also exist in spoken rather than written form. Incorporating audio input alongside textual data is crucial, as certain dialectal variations manifest at the phonotactic level. Last but not least, currently evaluation and promotion work on the interface is biased on the fact that most of the participants are highly-educated people and they at least hold a bachelor degree. In the future we are aiming for conducting more formal experiments on the local communities and adapt the interface to satisfy their needs and concerns.

## Ethical Considerations

In developing the Dia-Lingle platform, we have carefully considered ethical implications of dialectal data collection. Our interface content has been designed to exclude sensitive, discriminatory, or potentially offensive language. All text elements appearing throughout the interface have been reviewed and verified by native speakers to ensure cultural appropriateness and linguistic accuracy for most of the language options. Currently, we offer ten language options: English (the reference language), French, Italian and Spanish (machine-translated), and German, Simplified Chinese, Japanese, Korean, Latvian and Kurdish (machine-translated and subsequently verified by native speakers).

To prevent potential conflicts related to dialectal identity, the platform does not permit users to arbitrarily modify dialect names, which helps avoid contentious naming disputes that could arise from different sociolinguistic perspectives.

## Acknowledgments

We want to express our genuine gratitude to Eunjung Cho, Yuxuan Zhang, Raphaël Baur and Michitatsu Sato for externally helping us verify Korean, German and Japanese machine-translated texts in Dia-Lingle. In particular, Eunjung Cho and Michitatsu Sato have also helped us create the visualisation on Korean and Japanese dialects respectively. This work was supported by the Swiss National Science Foundation through the MUTAMUR project (no. 213976) and the Personalized Visual Analytics project (no. 10003068).

## References

- Sina Ahmadi, Daban Q. Jaff, Md Mahfuz Ibn Alam, and Antonios Anastasopoulos. 2024. [Language and speech technology for Central Kurdish varieties](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC/COLING 2024*, pages 10034–10045. ELRA and ICCL.
- Md Mahfuz Ibn Alam, Sina Ahmadi, and Antonios Anastasopoulos. 2024. [CODET: A benchmark for contrastive dialectal evaluation of machine translation](#). In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 1790–1859. Association for Computational Linguistics.
- Nicholas Asher, Julie Hunter, Mathieu Morey, Benamara Farah, and Stergos Afantenos. 2016. [Discourse structure and dialogue acts in multiparty dialogue: the STAC corpus](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 2721–2727, Portorož, Slovenia. European Language Resources Association (ELRA).
- Peter Auer. 2005. Europe’s sociolinguistic unity, or: A typology of European dialect/standard constellations. *Perspectives on variation: Sociolinguistic, historical, comparative*, 7:7–42.
- Jürgen Bernard, Marco Hutter, Matthias Zeppelzauer, Dieter Fellner, and Michael Sedlmair. 2018. [Comparing visual-interactive labeling with active learning: An experimental study](#). *IEEE Transactions on Visualization and Computer Graphics*, 24(1):298–308.
- ElMehdi Boujou, Hamza Chataoui, Abdellah El Mekki, Saad Benjelloun, Ikram Chairi, and Ismail Berrada. 2021. [An open access NLP dataset for arabic dialects: Data collection, labeling, and model construction](#). *CoRR*, abs/2102.11000.
- John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.*, 189.
- José Camacho-Collados, Kiamehr Rezaee, Talayeh Riahi, Asahi Ushio, Daniel Loureiro, Dimosthenis Antypas, Joanne Boisson, Luis Espinosa Anke, Fangyu Liu, and Eugenio Martínez Cámara. 2022. [TweetNLP: Cutting-edge natural language processing for social media](#). In *Proceedings of the The 2022 Conference on Empirical Methods in Natural Language Processing - System Demonstrations*, pages 38–49. Association for Computational Linguistics.
- Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. [From game design elements to gamefulness: defining “gamification”](#). In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, page 9–15. ACM.
- Daniel Deutsch, Eleftheria Briakou, Isaac Caswell, Mara Finkelstein, Rebecca Galor, Juraj Juraska, Geza Kovacs, Alison Lui, Ricardo Rei, Jason Riesa, et al. 2025. WMT24++: Expanding the language coverage of WMT24 to 55 languages & dialects. *arXiv preprint arXiv:2502.12404*.
- Pelin Dogan-Schönberger, Julian Mädler, and Thomas Hofmann. 2021. [SwissDial: Parallel multidialectal corpus of spoken Swiss German](#). *Preprint*, arXiv:2103.11401.
- Pierre Godard, Gilles Adda, Martine Adda-Decker, Juan Benjumea, Laurent Besacier, Jamison Cooper-Leavitt, Guy-Noel Kouarata, Lori Lamel, Hélène Maynard, Markus Mueller, Annie Rialland, Sebastian Stueker, François Yvon, and Marcely Zanon-Boito. 2018. [A very low resource language speech corpus for computational language documentation experiments](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. ELRA.
- Annika Grützner-Zahn, Federico Gaspari, Maria Gigakou, Stefanie Hegele, Andy Way, and Georg Rehm. 2024. [Surveying the technology support of languages](#). In *Proceedings of the Second International Workshop Towards Digital Language Equality (TDLE)*, pages 1–17. ELRA and ICCL.
- Barry Haddow, Rachel Bawden, Antonio Valerio Miceli Barone, Jindřich Helcl, and Alexandra Birch. 2022. [Survey of low-resource machine translation](#). *Computational Linguistics*, 48(3):673–732.
- Stuart Hallifax, Maximilian Altmeyer, Kristina Kölln, Maria Rauschenberger, and Lennart E. Nacke. 2023. [From points to progression: A scoping review of game elements in gamification research with a content analysis of 280 research papers](#). *Proc. ACM Hum.-Comput. Interact.*, 7(CHI PLAY).
- Aditya Joshi, Raj Dabre, Diptesh Kanojia, Zhuang Li, Haolan Zhan, Gholamreza Haffari, and Doris Dippold. 2024. [Natural language processing for dialects of a language: A survey](#). *CoRR*, abs/2401.05632.
- Yoolim Kim, Vita V Kogan, and Cong Zhang. 2023. [Collecting big data through citizen science: Gamification and game-based approaches to data collection in applied linguistics](#). *Applied Linguistics*, 45(1):198–205.



- Jeanine Krath, Linda Schürmann, and Harald F.O. von Korflesch. 2021. [Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning](#). *Computers in Human Behavior*, 125:106963.
- Thomas Lippincott and Ben Van Durme. 2021. [Active learning and negative evidence for language identification](#). In *Proceedings of the Second Workshop on Data Science with Human in the Loop: Language Advances*, pages 47–51, Online. Association for Computational Linguistics.
- Alexandre Magueresse, Vincent Carles, and Evan Heetderks. 2020. [Low-resource languages: A review of past work and future challenges](#). *Preprint*, arXiv:2006.07264.
- Ramesh Manuvinaurike and David Devault. 2015. [Pair me up: A web framework for crowd-sourced spoken dialogue collection](#). In *Natural Language Dialog Systems and Intelligent Assistants*, page 189–201.
- Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhersch, and Armand Joulin. 2018. [Advances in pre-training distributed word representations](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*.
- Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Taiwo Fagbohunge, Solomon Oluwale Akinola, Shamsuddeen Muhammad, Salomon Kabongo Kabenamualu, Salomey Osei, Freshia Sackey, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene Ahia, Musie Meressa Berhe, Mofetoluwa Adeyemi, Masabata Mokgesi-Seling, Lawrence Okegbemi, Laura Martinus, Kolawole Tajudeen, Kevin Degila, Kelechi Ogueji, Kathleen Siminyu, Julia Kreutzer, Jason Webster, Jamiil Toure Ali, Jade Abbott, Iroro Orife, Ignatius Ezeani, Idris Abdulkadir Dangana, Herman Kamper, Hady Elsahar, Goodness Duru, Ghollah Kioko, Murhabazi Espoir, Elan van Biljon, Daniel Whitenack, Christopher Onyefuluchi, Chris Chinenye Emezue, Bonaventure F. P. Dossou, Blessing Sibanda, Blessing Bassey, Ayodele Olabiya, Arshath Ramkilowan, Alp Öktem, Adewale Akinfaderin, and Abdallah Bashir. 2020. [Participatory research for low-resourced machine translation: A case study in African languages](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2144–2160, Online. Association for Computational Linguistics.
- Haruna Ogawa, Hitoshi Nishikawa, Takenobu Tokunaga, and Hikaru Yokono. 2020. [Gamification platform for collecting task-oriented dialogue data](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 7084–7093, Marseille, France. European Language Resources Association.
- Fredrik Olsson. 2009. A literature survey of active machine learning in the context of natural language processing. page 59.
- Kyubyong Park, Yo Joong Choe, and Jiyeon Ham. 2020. [Jejeuo datasets for machine translation and speech synthesis](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2615–2621. ELRA.
- Massimo Poesio, Jon Chamberlain, Udo Kruschwitz, Livio Robaldo, and Luca Ducceschi. 2013. [Phrase detectives: Utilizing collective intelligence for internet-scale language resource creation](#). *ACM Trans. Interact. Intell. Syst.*, 3(1).
- Yves Scherrer and Owen Rambow. 2010. Natural language processing for the Swiss German dialect area. In *Semantic Approaches in Natural Language Processing: Proceedings of the 10th Conference on Natural Language Processing, KONVENS 2010*, pages 93–102. Universitätsverlag des Saarlandes.
- Katie Seaborn and Deborah I. Fels. 2015. [Gamification in theory and action: A survey](#). *International Journal of Human-Computer Studies*, 74:14–31.
- Erin Snyder and Jason R Hartig. 2013. [Gamification of board review: a residency curricular innovation](#). *Medical Education*, 47(5):524–525.
- Shinnosuke Takamichi and Hiroshi Saruwatari. 2018. [CPJD corpus: Crowdsourced parallel speech corpus of Japanese dialects](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Armando M. Toda, Wilk Oliveira, Ana C. Klock, Paula T. Palomino, Marcelo Pimenta, Isabela Gasparini, Lei Shi, Ig Bittencourt, Seiji Isotani, and Alexandra I. Cristea. 2019. [A taxonomy of game elements for gamification in educational contexts: Proposal and evaluation](#). In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, volume 2161-377X, pages 84–88.
- Gustavo F. Tondello, Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. 2016. [The gamification user types hexad scale](#). In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play, CHI PLAY '16*, page 229–243, New York, NY, USA. Association for Computing Machinery.
- Marcos Zampieri, Preslav Nakov, and Yves Scherrer. 2020. [Natural language processing for similar languages, varieties, and dialects: A survey](#). *Natural Language Engineering*, 26(6):595–612.
- G. Zichermann and C. Cunningham. 2011. [Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps](#). O'Reilly Series. O'Reilly Media.
- Rob Zwitterlood, Marjan ter Harmsel, Johanna Schulting, Karin Wiefferink, and Ellen Gerrits. 2022. [To game or not to game? efficacy of using tablet games in vocabulary intervention for children with DLD](#). *Applied Sciences*, 12(3).

## A Appendix

In this section, we include the detailed overview of the Review stage from the Dia-Lingle interface as shown in Figure 7. We also include the content of the survey questionnaire in Table 1 and instructions for the usability studies in Table 2. Usability studies were conducted with a total of four participants. The strengths reported from participants are displayed in Table 3 and the problems identified during these studies are presented in Table 4.

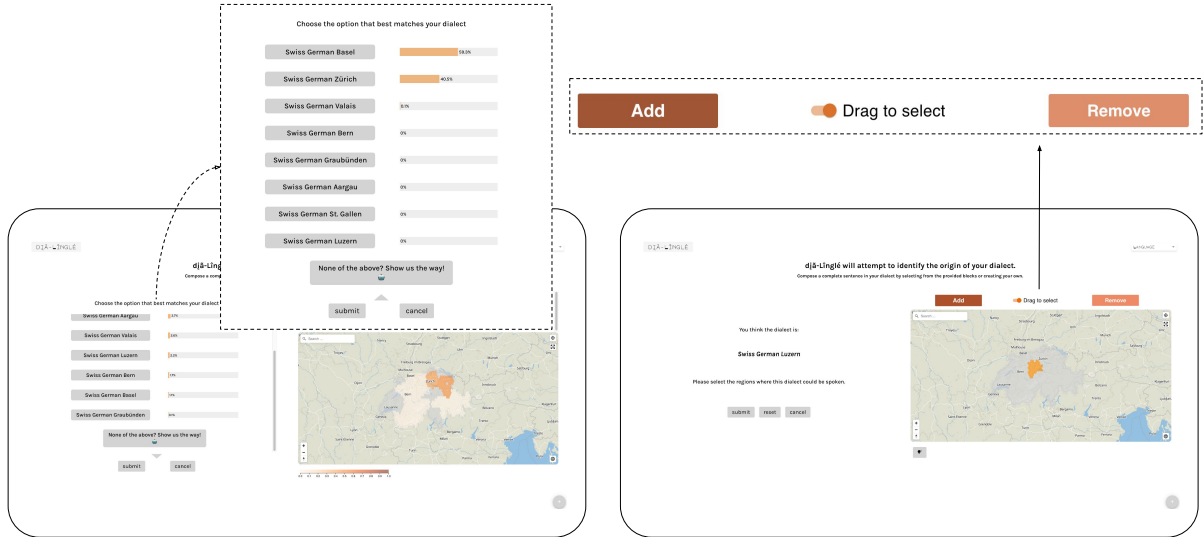


Figure 7: Overview of the Dia-Lingle interface design at the Review stage, detailed in Section 4.4. The interface displays a list of all possible dialects and offers a lasso tool enabling users to define more accurate dialect regions.

Section	Question
Demographic Questions	<ol style="list-style-type: none"> <li>How old are you?</li> <li>What is the highest level of education you have completed?</li> <li>What is your native language?</li> <li>Which language did you select as the theme of the game (i.e., the language whose dialectal variations you explored)?</li> <li>How familiar are you with the language you explored in the game?</li> </ol>
SUS Questions	<ol style="list-style-type: none"> <li>I think I would use this interface frequently.</li> <li>I found the interface unnecessarily complex.</li> <li>I thought the interface was easy to use.</li> <li>I think that I would need the support of a technical person to be able to use this interface.</li> <li>I found the various functions in this interface were well integrated.</li> <li>I thought there was too much inconsistency in this interface.</li> <li>I would imagine that most people would learn to use this interface very quickly.</li> <li>I found this interface very cumbersome to use.</li> <li>I felt very confident using the interface.</li> <li>I needed to learn a lot of things before I could get going with this interface.</li> </ol>
Feedback	<ol style="list-style-type: none"> <li>What do you like the most about Dia-Lingle?</li> <li>What do you dislike the most about Dia-Lingle?</li> <li>ask for more feedback &amp; contributions.</li> </ol>

Table 1: Detailed content of the survey distributed for usability study.

Stages	Sub-stages	Duration	Description
Stage I	Welcome Stage	1 min	Begin by briefly introducing yourself
Stage II	Stage A	2 mins	Semi-Structured Interview for Expectation Check
	Stage B	2 - 3 mins	Explain motivations + UI design
Stage III	Stage C	10 - 15 mins	Participants try interacting with the interface on Quiz Game and/or Match Game
Stage IV	Break Stage	1 min	A short break
Stage V	Stage D	3 - 5 mins	Explain models + design choices
	Stage E	2 - 4 mins	Concluding Semi-Structured Interview
Stage VI	Stage F	2 mins	Informal Questionnaire
Stage VII	End Stage	1 min	End the study and thank the participant

Table 2: Table of instructions of Usability Studies. The complete duration of the study is expected to be approximately 30 minutes. At the sub-stage F, we hand out the same survey shown in Table 1 to the participants.

Strengths reported in Usability Studies	P1					P2					P3					P4				
	E	C	Q	R	M	E	C	Q	R	M	E	C	Q	R	M	E	C	Q	R	M
multilingual support																				
rolling pins to catch attention																				
interface is very transparent																				
animation of component transition																				
hovering effects																				
pop-up window to prevent from accidental click																				
smooth transition of components																				
typing suggestions																				
dialect visualisation on map																				
gamified difficulty levels																				
onboarding guidelines for lasso tool																				
classifier is relatively robust																				
the overall idea																				

Table 3: A detailed distribution matrix of usability strengths showing which spotlights found in which step by which participant in Usability Studies. **P1; P2; P3; P4** represent four participants. **E; C; Q; R** and **M** represent different stages of *Entry*; *Choice*; *Quiz*; *Review* and *Match* on dashboard detailed in Section 4.

Problems reported in Usability Studies	P1					P2					P3					P4				
	E	C	Q	R	M	E	C	Q	R	M	E	C	Q	R	M	E	C	Q	R	M
some components are small																				
unclear about how to proceed																				
feel <i>Match</i> is harder than <i>Quiz</i>																				
click the components before onboarding guidelines finish																				
do not like the text input blocks																				
do not understand colour legend																				
do not understand probability																				
want more explanations on prediction																				
yellow ticks are confusing																				
do not understand the standardised sentence																				
want audio input option																				
want instruction texts larger																				

Table 4: A detailed distribution matrix of usability problems showing which problems found in which step by which participant in Usability Studies. **P1; P2; P3; P4** represent four participants. **E; C; Q; R** and **M** represent different stages of *Entry*; *Choice*; *Quiz*; *Review* and *Match* on dashboard detailed in Section 4.