Rěc :: A Language Learning System

Lilli-Chiara Kurth

27.03.2024
Universität der Künste Berlin
Technische Universität Berlin
M.A. Design & Computation
lilli.chiaraa@googlemail.com

Abstract

This project aims to contribute to the efforts of preserving the sorbian culture by creating a language learning system, that not only makes the lower sorbian language more easily accessible, but serves as a (first) point of contact with the sorbian culture as well. The overarching goal for this system is to be useful, enticing and fun for language learners of all ages above 10. To achieve this, the teaching strategy is based on narration, specifically of stories and traditions embedded in the sorbian culture. The visual identity, as well as the base material for the stories, are created and curated in cooperation and collaboration with sorbian artists. By actively involving the sorbian community, I expect the system and its content to be as authentic as possible to the current sorbian culture, presented in a playful way that sparks further interest in the sorbian people. With this approach, Rěc *is an introduction to the sorbian culture, as well as an exploration in digital creative learning. Various optimization techniques are employed to create an effective structure and rhythm of the learning experience, while close contact and collaboration with sorbian native speakers ensures the authenticity and correctness of the content.*

Figure 1: *Rěc homepage*

I. The Problem

Die Sorben, the sorbish people, are a slavic ethnic group that settled in the east of germany around 1,500 years ago. [1] They are recognized as a national minority and connected by specific traditions, clothing, their own languages: *Upper Sorbian*, *Lower Sorbian*, flag and national anthem. Since sorbian people usually have a german citizenship and there is no sorbian nation, it is a wicked problem to keep sorbian traditions and especially the languages alive. [11]

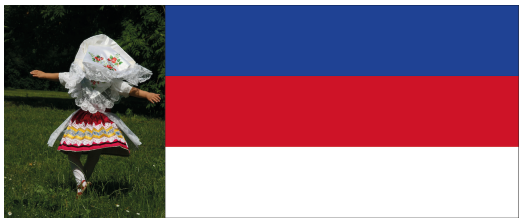


Figure 2: *Traditional clothing and national flag*

I want to take part in saving the sorbian culture because I grew up in the sorbian area and consider myself to be half sorbian. I participated in traditional festivities, learned the lower sorbian language and experienced first hand how difficult it is for sorbian families to not lose their cultural heritage. I myself struggled to find a sorbian community after moving from home to pursue my studies, which is why I lost my language abilities and connections to the sorbian way of life. *Rěc* can be considered as an expression of my longing for a convenient point of access to the sorbian world, as well as my desire as a computational linguist and creative writer to investigate narration as the main tool for teaching. This approach to the project is the reason why I chose *Accessibility & Convenience*, *Efficiency & Correctness* and *Authenticity & Excitement* as the basic design objectives for my system. I emphasize the involvement of sorbian native speakers and artists because I want the system to be built by the community for the community. See appendix A for a problem solution mind map.

I.I. Accessibility & Convenience

Currently, there are only around 60,000 sorbian people [11] living in a specific east german area [3], which creates a locally exclusive point of access to the sorbian culture. I design and build my system such that people are enabled to learn the language without the need for being physically present with a sorbian teacher. This will benefit people leaving the sorbian area for places where no sorbian is spoken, people who are interested in the language but are unable to connect with sorbian teachers, or students who want to explore and learn the language on their own, additionally to what is presented to them in class. It might also be used by sorbian teachers to enrich their teaching. The overarching goal is to create an easily accessible space for learning lower sorbian. This is why the system needs to be digital. Conveniently, I am able to integrate my system with the knowledge graph back-end built by Prof. Dr. Daniel D. Hromada in the context of his project *Digital Primer Implementation of Human-Machine Peer Learning*. [4] This frees me to focus on involving the community and building the teaching content of *Rěc*, while the fully functional technical setup by Hromada allows for instant realization.

I.II. Efficiency & Correctness

Rěc is meant to teach linguistically correct content in an effective manner, which demands a sensible teaching strategy, a fitting tool structure and extensive knowledge about the language. This becomes even more important since the content is presented in an unusual way through narration. To realize efficiency and correctness, I have analyzed the lower sorbian language itself and the highly developed language learning apps Duolingo and Mimo. This provided me with insights about the demands of learning lower sorbian, as well as about tool features that can be utilized for optimizing various levels of a digital learning system. Employing this information facilitates an efficient way of teaching the sorbian alphabet and its special letters, phonetics of the special letters, as well as grammar and vocabulary.

I.III. Authenticity & Excitement

Aside from being efficient and correct, the learning process through the system is meant to be fun and spark interest in the sorbian culture. This is realized by a strong visual identity, gamification and employing narration as the main form of teaching. The linguistic content of the learning units is presented in form of stories that are centered around sorbian traditions (e.g. Hahnrupfen *Łapanje kokota*, Vogelhochzeit *Ptaškowa swajźba* and legends (e.g. *Lutkis, Krabat*), and visualized in collaboration with sorbian artists. This community-involving approach will result in a visually unique, very authentic impression of the current sorbian culture. Hopefully, encountering the sorbian language, stories and traditions in this way will help spread awareness and enthusiasm for this minority, help keeping the languages alive and strengthen the sorbian community in the future.

II. Construction Methods

I employ multiple methods for ensuring the *linguistic correctness* of the content, finding an *effective teaching strategy*, and offering an *exciting learning experience*, thereby constructing an overall valuable language learning system, that adheres to the basic design objectives described above.

II.I. Linguistic Correctness: studying lower sorbian

Finding extensive and reliable resources about lower sorbian, its structure and how it works, is crucial for ensuring that the content of the teaching is linguistically correct. I rely on three main sources for obtaining such information.

Dr. Gerald Stone is one of the main contributors to linguistic research regarding upper and lower sorbian: *I seem to be the only person in the world whose native language is English, who can understand nineteenth-century Upper Sorbian printed in Gothic type.* [16] He has a long list of publications about various topics regarding both sorbian languages. I reference his chapter *Sorbian (Upper and Lower)* [17] in my work for this project.

The Sorbian Institute is officially responsible for all efforts regarding folklore research and academic work in the fields of cultural studies and linguistics. [6] I am in close contact with them and am verifying the systems content through them. They have been very helpful and forthcoming with supporting the project and providing data already. I will continue to work with them closely.

In addition to the more scientific sources of language knowledge, I rely on conversing with various members of the sorbian community. A casual perspective on the current usage of lower sorbian in daily life is very valuable and will shape the teaching as well.

II.II. Effective Teaching Strategy: popular language learning apps

Since *Rěc* is a digital system designed to be fully functional without the presence of any human teachers, it shares certain demands and characteristics with popular language learning apps, namely Duolingo and Mimo. I have implemented certain features I discovered in my analysis of the two apps to optimize *Rěc*.

Duolingo is with 72,6 mio active monthly users in march 2023 [14] one of the most popular language learning apps on the market. They have developed a multi-layered approach for teaching by employing various learning strategies, while focusing on useful and actionable content. They allow for implicit learning, let learners encounter *unexpected* and therefor memorable content and offer *interactive formats* to include all types of language learners. [2] Additionally, they foster a layer of competition by emphasizing and rewarding daily activity by a specific feature; *the streak*. In combination, all of those strategies lead to the success people experience when learning and practicing a language with the app.

Mimo is with 10 mio total downloads [10] less popular than Duolingo, but also caters to a very different user base. Mimo is designed to teach programming languages like Python or Apples language Swift. They are a diverse group of teachers, engineers, strategists, and designers that work on finding the best way to make coding accessible to everyone. [10] In their efforts of teaching not-quite-human languages, they have developed very interesting similarities to Duolingos approach to teaching: They also implement a streak feature, the learning units are *structured for daily practice* and are *organized to comprehensive learning paths*. The exercises are designed interactively as well, there are rewards for consistent learning and the content is *gamified*.

II.III. Exciting Learning Experience: narration & gamification

I am combining narration as the main form of teaching with elements of gamification and the insights from analyzing Duolingo and Mimo, to create a very unique and exciting learning experience. I expect this way of teaching to be more fun and effective than relying on traditional teaching methods like repetition and gap texts only. This is partly verified by the success of Duolingo and Mimo. Additionally, I will be conducting a survey with beta-testers of the system in the next iteration of this project to further investigate the nature and effectiveness of the experience provided by the system.

The narrative structure is inspired by *Alice in Wonderland* written by Lewis Carroll in 1865: There is a main character: Žaba *Frosch*, exploring a wonderful world, encountering strange beings and having fantastical adventures. The beings and adventures are taken from or inspired by sorbian folklore and traditions. Each story/ tradition I pick to be implemented in the system is transformed to be structured as a learning path. Each path has multiple learning units and knowledge tests, teaching a specific aspect of the sorbian language by narrating little situations featuring Žaba and the current story/ tradition. A visual representation of this approach can be found in appendix B. The exemplary learning path narrating the sorbian celebration of the beginning of spring *Ptaškowa swajźba*, is constructed to teach the special letters of the sorbian alphabet by Žaba interacting with various guests at the bird wedding. This is the initial learning path and the first story a learner encounters. I will decide on what further stories to use depending on what works well regarding the teaching objective of the future learning paths. Since I plan on visualizing the stories in collaboration with sorbian artists, I will make sure to let them pick their favourite stories to work on. This will influence my selection of stories as well.

Gamification is used strategically to accompany the narrative structure of the content and serves to optimize various aspects of the learning experience. I talk about this in detail in the next chapter. Overall, they add layers of playfulness, competition and motivation to the system, which is meant to result in a fun, enticing and effective teaching situation learners want to come back to.

III. Optimization

Following the basic design objectives and construction methods described above, I have implemented multiple optimization techniques to support the optimal development of my system, as well as the learning experience itself. I optimize *Rěc* on the levels of *Teaching Efficiency, Inviting Consistent Activity, Content Memorability* and *Scalability*. Most of the system features and design decisions resulting from my optimization efforts are gamifying by nature, adding play, competition and motivative incentives to the system. In combination with the teaching being based on narration, this has resulted in *Rěc* to be set up and feel more like a game than a learning system. Additionally, the optimization process, specifically ensuring *Scalability*, served as proof of concept for my ability to realize certain system behaviors and to upkeep them long term.

III.I. Teaching Efficiency

The structure and teaching rhythm of *Rěc* is carefully chosen to enable a fruitful learning experience. Presenting the content in a *self paced* manner gives learners enough time to comprehend, form connections and discover patterns, which allows for implicit and strategic learning. By simultaneously prompting the learner to *interact* with the system before the next chunk of content can be seen, I ensure they actively use the language on multiple levels (writing, speaking, comprehending). Lastly, I strategically identified a valuable *initial vocabulary basis* to teach.

Self Paced Learning is shaping the basic rhythm of encountering content within the system. The idea is to let the learner progress through the lections and stories with their own speed. This is made possible by capsulating each learning path within their own story, complete in themself. It will be possible to unlock and follow multiple learning paths, stories, at the same time, and to choose what learning unit from what learning path will be solved today. This will be limited by ranking the paths/ stories in difficulty, only letting them become available to the learner when they have progressed to a certain level of language mastery. The learning units themself are designed to be self paced as well. Each sentence is presented by itself. Its written in german and sorbian, with an audio sample of the sorbian sentence, recorded by a native speaker. A visual example of this can be seen in Figure 3 below. The next sentence is shown only when the learner presses a button or clicks with their mouse. This slows the learning situation to a pleasant pace and keeps the learner from being overwhelmed with all of the information at once. Focusing the attention on very small chunks of input at a time frees the learner to not only comprehend the material fully, but to "read between the lines" as well. Still, each learning unit is supposed to be fully comprehensible within around 10min.

Interactivity is the second feature shaping the rhythm of encountering content within the system. Depending on what a learning unit is supposed to teach, the learner is prompted to interact with *Rěc*: they might need to say a specific word correctly, spell something, pick a fitting word or sentence etc. before the next sentence can be revealed. Actively using the language in multiple ways is crucial for developing communication skills on all levels a language operates on. Since the demands of interaction are limited to singular words or sentences at a time by the self paced learning approach, the playful nature of the content setup will not be interrupted. Naturally, the knowledge tests concluding each learning path are fully interactive as well, since they are supposed to challenge the learners language abilities. I will specifically test for spelling, articulation/ pronunciation and comprehension.

Finding the **Initial Vocabulary Basis** to teach turned out to be a complex process. Since the system is supposed to teach a useful and comprehensible vocabulary basis, I needed to contend with the question of how many and which words of the lower sorbian language exactly to focus on first in teaching. The dictionary available through the Sorbian Institute [7] is described to contain 82,000 entries, which indicates that there are around 82,000 more or less relevant words in the lower sorbian language I needed to consider and rank for their relevancy somehow, to find the optimal initial vocabulary basis. I decided to identify a manageable set of around 200 words, that will allow the learner to produce the greatest amount of useful utterances for casual conversation. How exactly this optimization objective brought me to the final version of the list is described in chapter IV: Data, since this exceeds the scope of this paragraph. This final list is the initial vocabulary basis for the system and will be used for developing the *Weekly Vocabulary Challenges*. I expect the learner to pick up on additional words while encountering the learning unit sentences, but this initial vocabulary basis is chosen to be intentionally teached.

III.II. Inviting Consistent Activity

Rewarding consistency and creating something to look forward to, to strive for, to log on for, is crucial for the system to be desirable enough to consistently come back to.

The Streak Feature is one of the most motivating components of both language learning apps [5], according to my research. Aside from wanting to progress in the story, I expect building the streak to be one of the main reasons for consistent daily activity. To realize an enticing streak system with rewards, I have designed the visual interface of *Rěc* such that the image of the main character Žaba is interchangeable. By reaching certain streaks (e.g. solving 1 learning unit per day 10/20/30/... days in a row) a new main character image is unlocked, which can then be chosen by the learner and displayed in future learning units. A selection of main character images can be found in appendix C.

Figure 3: *Žaba, first learning unit*

I expect the ability to customize the visual appearance of the system by solving just 1 learning unit per day to be very motivating long term, and I expect the same for the wonder of how the next main character image may look like. Simultaneously, the effort spent to actually be active daily for long periods of time becomes visible with a comprehensible number: a high day count is very tangible, which is why the streak is prominently displayed on the homepage (see Figure 1). This setup adds a layer of validation to the learning experience, further optimizing the motivating factor of *Rěc* by enticing and rewarding daily practice.

The Weekly Vocabulary Challenge requires the learner to master a curated set of around 6 sorbian words by the end of 7 days. This will be tested and rewarded with a "cheat day": in the monthly view of the streak (see appendix E), a day of inactivity can be healed such that a lost streak might be saved. This is a strong incentive for the learner to actually do the work and learn new words, since it will enable them to "cheat" or "strategically pause" the learning routine for up to 4 days per month. I expect this to be very motivating for learners using the system daily over long periods of time, since they are more likely to be able to uphold their streak despite the ups and downs of life. How I curate the sets of words to learn during the challenges is described in chapter IV: Data as well, since they are based on the initial vocabulary basis.

In combination, the daily streak and the weekly vocabulary challenge provide short term rewards and long term motivation for solving learning units and consequently for practicing consistently. At the same time, they add layers of customization, competition and validation to the system.

III.III. Content Memorability

Optimizing the system for memorability serves two intentions: enticing attention and aiding the learning process by the material being easy to stay in mind. *Unexpected Encounters* provide the learner with exciting and therefor easily remembered material. A *Strong Visual Identity* with a high level of quirkiness is very striking, hopefully enjoyable and easy to freely come back to.

The feature I call **Unexpected Encounters: ODD PICS** is resulting from discussing my project with Max Baraitser Smith, who at this time was working on another language-related project: *Creating Maps Using Free Word Association Games*. [15] He voiced the idea of coupling the learning experience with an odd activity to make it more memorable. One could imagine

going to the park for learning words like *grass, bench, tree, walking, squirrel*, and by actually sitting on a bench and smelling the grass, the learning is imagined to be more tangible, multi-layered and long-lasting. I resonated with that theory, and discovering that Duolingo strategically uses *quirky sentences* to *push learners to think carefully about the language they're learning* [2], has verified the value of following this approach of utilizing oddness. I decided to accompany each *Weekly Vocabulary Challenge* with its own ODD PIC, an image generated with Stable Diffusion, depicting all or most of the words to be learned. An exemplary ODD PIC can be seen in appendix F. With the addition of audio samples recorded by sorbian native speakers, this is a playful but useful learning basis, enabling the learner to actually win the challenge. The OOD PIC is easily printable, in case a physical object is wanted, and can be referenced any time within the system. Since the ODD PIC is very striking, it is supposed to aid the memory of the learner.

I expect a **Strong Visual Identity** to support a strong learning process by enticing attention and creating wonder and enjoyment. Since I believe my future userbase to be primarily highschool students, I chose to develop a colorful, loud and quirky visual appearance, mirroring the quirkiness of the sorbian folklore and following the current campy internet aesthetic. I plan on collaborating with various sorbian artists for visually accompanying the learning units, paths and stories, which will result in a very diverse set of content illustrations. This is part of the goal for my system to authentically reflect the current sorbian culture. Aside from this, the visual content for *The Streak Feature* and the *Weekly Vocabulary Challenge* is generated with Stable Diffusion. Contrasting the human input, the generated images add to the visual quirkiness I am striving for, while also benefiting *Scalability*. The current visual state of *Rěc* can be considered as prototype I, see Figure 1 and appendix C, D, E. By pushing for a unique visual appearance, I hope to set my system apart from other digital environments.

III.IV. Scalability

Being able to comfortably scale the system as needed, adds to the convenience of *Rěc* and is very important for being able to uphold the system long term, support multiple users with various needs and expand the teaching content in the future, while keeping the system fully functional. This is made possible by involving multiple sorbian artists and *Stable Diffusion* for creating and generating visual content, and by structuring the learning material in a *compartmentalized* manner.

The learning material is highly **compartmentalized**, which results from the *Self Paced Learning* approach and allows for a comfortable scaling rhythm. Each learning path covers its own story, its own teaching objective, its own exercises and knowledge tests, thereby being concluded in themself. This allows me to incrementally expand the content of my system story by story, which ensures that *Rěc* is usable during further development. It also simplifies the involvement of the community, since each artist can focus on visualizing a single story/ learning path, allowing for parallel development and minimal dependency.

I am mostly relying on **Stable Diffusion**, a large text to image model able to produce state of the art results for image generation [12], to produce the visual content needed for *The Streak Feature*: main character images and the *Weekly Vocabulary Challenge*: ODD PICs. I am currently planning on unveiling a new character image every 10 streak days, which means I would need 36 main character images to cover a streak of 1 year. I have already generated well over 100 suitable images, enabling me to quickly update the system if needed. Following the identification of the *Initial Vocabulary Basis*, which is a set of around 132 words, I need to create around 22 ODD Pics to cover all *Weekly Vocabulary Challenges* resulting from this basis. With Stable Diffusion this is not only manageable, but also a fun process.

By using the generative power of Stable Diffusion and working compartmentalized with multiple artists, I am able to scale the visual content production with minimal effort depending on my system needs, as well as update the learning material incrementally.

IV. Data

I am working primarily with text-based data: sorbian literature and my own writings. I have access to sorbian text corpora [8] developed by the Sorbian Institute and will record audio samples of sorbian native speakers, as well as use story visualizations from sorbian artists. I will also document any conversations with the people from the sorbian community I work with in the context of this project. Additionally, I create images with Stable Diffusion for the main character related to the streak feature, as well as for the OOD PICS serving as visual cues in the *Weekly Vocabulary Challenges*.

IV.I. Initial Vocabulary Basis

This analysis was an attempt of finding an answer to the question "What do I want to teach with *Rěc*, and what do I want to teach first?". This is related to the vocabulary, as well as the grammatical behavior of lower sorbian. My goal was to identify a manageable set of around 200 words, that will allow the learner to produce the greatest amount of useful utterances for casual conversation. The results of this investigation allow me to implement the *Weekly Vocabulary Challenges*, as well as decide on the teaching objectives of the first learning units I will develop in the next iteration of this project.

Starting my investigation, I have conducted **Frequency Analyses** with the help of the Sorbian Institute. They are based on three lower sorbian text corpora, described in Table 1 (see appendix G), covering the years 1990-2022 with a combined token volume of over 15mio. [8] For each corpus, the Institute has provided me with a sorted list of the 500 most frequent tokens. I computed the common elements of all three lists, which resulted in a list of 132 tokens. The words from this list are the most frequently used words in sorbian literature, from 1990 until 2022. Following their identification, I implemented a script using the lower sorbian - german dictionary [9] in order to translate them automatically. A screenshot from this list can be seen in appendix G as well, Figure 4. Most of the words are prepositions or pronouns and verbs of perception or movement. There are some nouns present as well, related to daily life in urban environments. So, according to frequency, being able to describe spacial relationships, movement and action, as well as human/ family/ emotional relationships is most frequently useful. Aside from this, the list is not very well suited to form the basis for my approach to vocabulary learning. Functional words like prepositions are learned much better if they are properly contextualized instead of being presented by themself and an ODD PIC.

Following these insights, I strategically contemplated on what would be sensible to teach in my stories, and what to teach with the *Weekly Vocabulary Challenges*. I decided on developing the following learning paths: special letters, grammatical gender, negation, prepositions, pronouns, questions, greetings, words regarding family & relationships, counting & numbers, week days & time-related words, colors. The *Weekly Vocabulary Challenges* will complement the teaching of these paths. They will feature nouns related to the paths contents and to daily life in urban environments overall, verbs of perception & movement, and adjectives.

Overall, I decided to abandon the idea of finding a fixed set of words of around 200 useful words to split into groups and then use for curating the *Weekly Vocabulary Challenge*. Instead, I will develop the challenges simultaneously to developing the learning paths. This will allow me to pick words that would actually be useful to know, while not being prominent

in the learning units. Since the learning units follow a daily rhythm, while the challenges follow a 7-day-rhythm, this approach will also ensure that the challenges follow the learning progress, such that no knowledge about greetings will be asked of the learner before they had a chance to solve the related learning units.

IV.II. Audio Samples

I will record audio samples of sorbian native speakers to present learners with authentic language examples. This might be single words as well as whole sentences. By preferring female native speakers for recording sorbian articulations, I want to address the lack of female voices in lower sorbian audio datasets [19] to create a more balanced and realistic representation of the language community. I suspect that this lack of female voices is true for sorbian literature as well. To combat both imbalances, I will collaborate mainly with female native speakers, as well as prefer female authors as much as possible for developing the learning units.

IV.III. Visual content

I will use visual content created by sorbian artists to illustrate the learning and create a strong visual identity for my system. I will credit the artists within the learning paths themself. All content I generate with Stable Diffusion or draw by myself will not be credited separately.

V. Results

With my theoretic and linguistic research I was able to strategically decide on my systems basic design objectives, functionalities and features, its visual identity, learning rhythm, teaching strategy, content and validation/ motivation system. I have proof of concept that I will be able to build my system as such, support it long term and scale it as needed. I have also started to work with the Sorbian Institute and am getting in contact with sorbian artists.

Overall, I consider the current state of my system to be a successful first prototype, ready to be properly realized.

VI. Next Steps

The very fist problem to solve next is the financial level of collaborating with people. I want to be able to pay them for their efforts, since translations, recordings and conversations about sorbian stories are very time-consuming and demanding. I have applied for funding to solve this problem and am eager to hear from them. Most of my efforts in the next iteration of the project will be focused on developing the learning paths and *Weekly Vocabulary Challenges*. Hopefully, this can be concluded with the visual input from sorbian artists. At some point - probably when the first 3 learning paths are implemented - I will present the system to beta-users and conduct a survey regarding the learning experience and teaching effectiveness. This will be done to verify how successful the teaching actually is and to further optimize my system. Lastly, I am currently testing *Twine*, an open-source tool for telling interactive, nonlinear stories. [18] This might be a great tool for developing the learning paths narratively.

VII. Acknowledgments

I am very grateful for the continuous support of Prof. Dr. Daniel D Hromada in realizing my vision for *Rěc*. I also want to thank the Sorbian Institute, specifically Fabian Kaulfürst and Madlen Domaschke, for their contributions in my development process and I am equally excited about the help of my friends connecting me with people from the sorbian community. Last but not least, my appreciation extends to the students of Design & Computation [13], who have created an environment of collaboration, constructive exchange and moral support while searching for ”Wicked Solutions“. This dynamic has enriched my project and the experience of developing it immensely.

References

[1] Aussiedlerbeauftragte. https://www.aussiedlerbeauftragte.de/Webs/AUSB/DE/themen/minderheiten-sprachgruppen/sorbisches-volk/sorbisches-volk.html. accessed: 2024-03-19.

[2] Duolingo. https://en.duolingo.com/approach. accessed: 2024-03-19.

[3] Förderverein Lausitz e.V. https://lausitz.de/de/kultur/sorben-wenden.html. accessed: 2024-03-21.

[4] Daniel Devatman Hromada and Hyungjoong Kim. Digital primer implementation of human-machine peer learning for reading acquisition: Introducing curriculum 2. 2023.

[5] Duy Huynh, Long Zuo, and Hiroyuki Iida. An assessment of game elements in language-learning platform duolingo. In *2018 4th International Conference on Computer and Information Sciences (ICCOINS)*, pages 1–4, 2018.

[6] Sorbisches Institut. https://www.serbski-institut.de/institut/. accessed: 2024-03-19.

[7] Sorbisches Institut. https://niedersorbisch.de/dnw/. accessed: 2024-03-19.

[8] Sorbisches Institut. https://niedersorbisch.de/korpus/komfort/zredla/. accessed: 2024-03-19.

[9] Sorbisches Institut. https://niedersorbisch.de/ndw/. accessed: 2024-03-19.

[10] Mimo. https://mimo.org/about-us. accessed: 2024-03-19.

[11] Minderheitensekretariat. https://www.minderheitensekretariat.de/die-lausitzer-sorben/. accessed: 2024-03-19.

[12] Onkar Mishra. https://medium.com/@onkarmishra/stable-diffusion-explained-1f101284484d. accessed: 2024-03-23.

[13] New Practice. https://www.newpractice.net/people. accessed: 2024-03-26.

[14] Duolingo Press Release. https://investors.duolingo.com/static-files/3c5dd359-4936-452a-9d9f-9a8c853c06c0. accessed: 2024-03-19.

[15] Max Baraitser Smith. Creating maps using free word association games. *Wicked Solutions*, 0, 2024.

[16] Dr. Gerald Stone. https://wendishresearch.org/category/gerald-stone/. accessed: 2024-03-19.

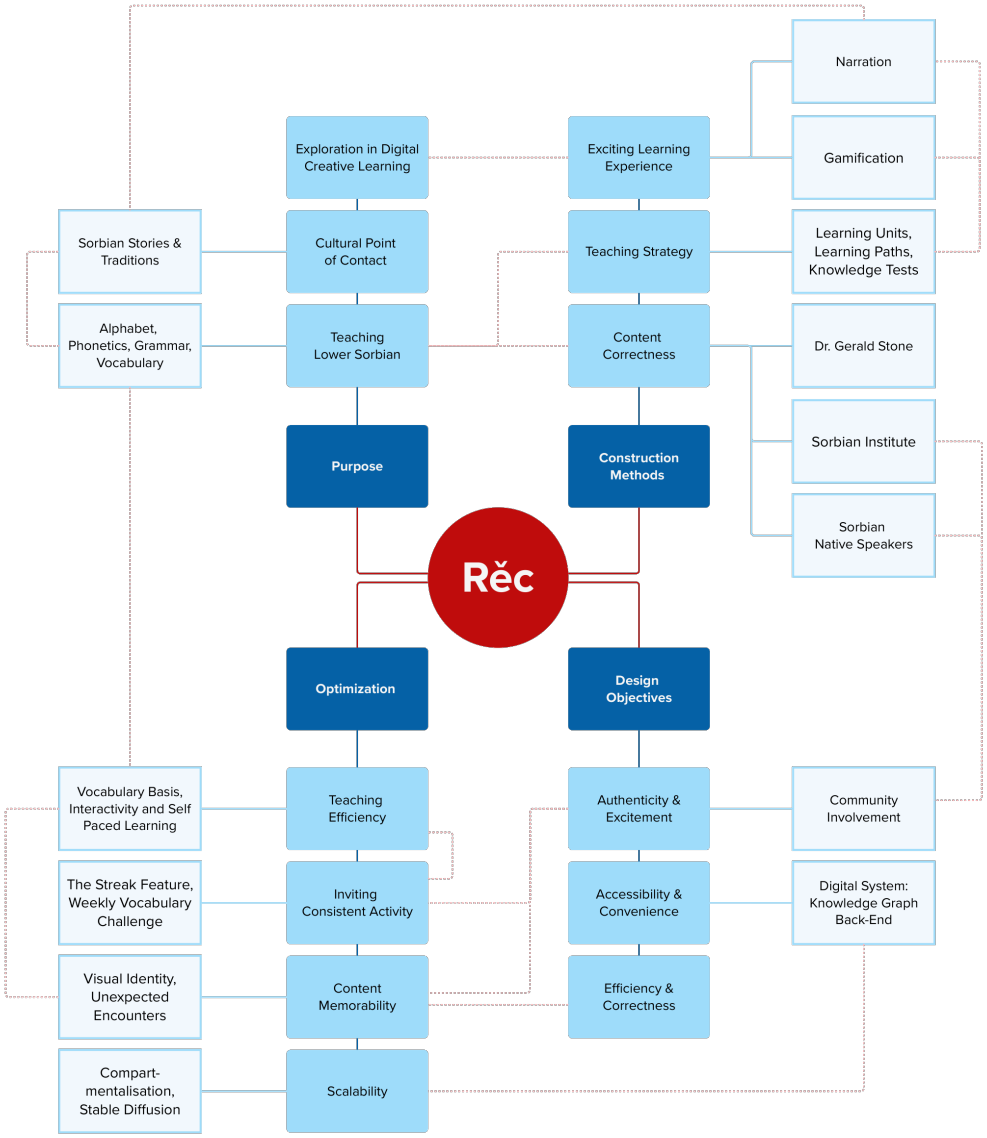
[17] Dr. Gerald Stone. *Sorbian (Upper and Lower)*. London New York: Routledge, 1993.

[18] Twine. https://twinery.org/. accessed: 2024-03-25.

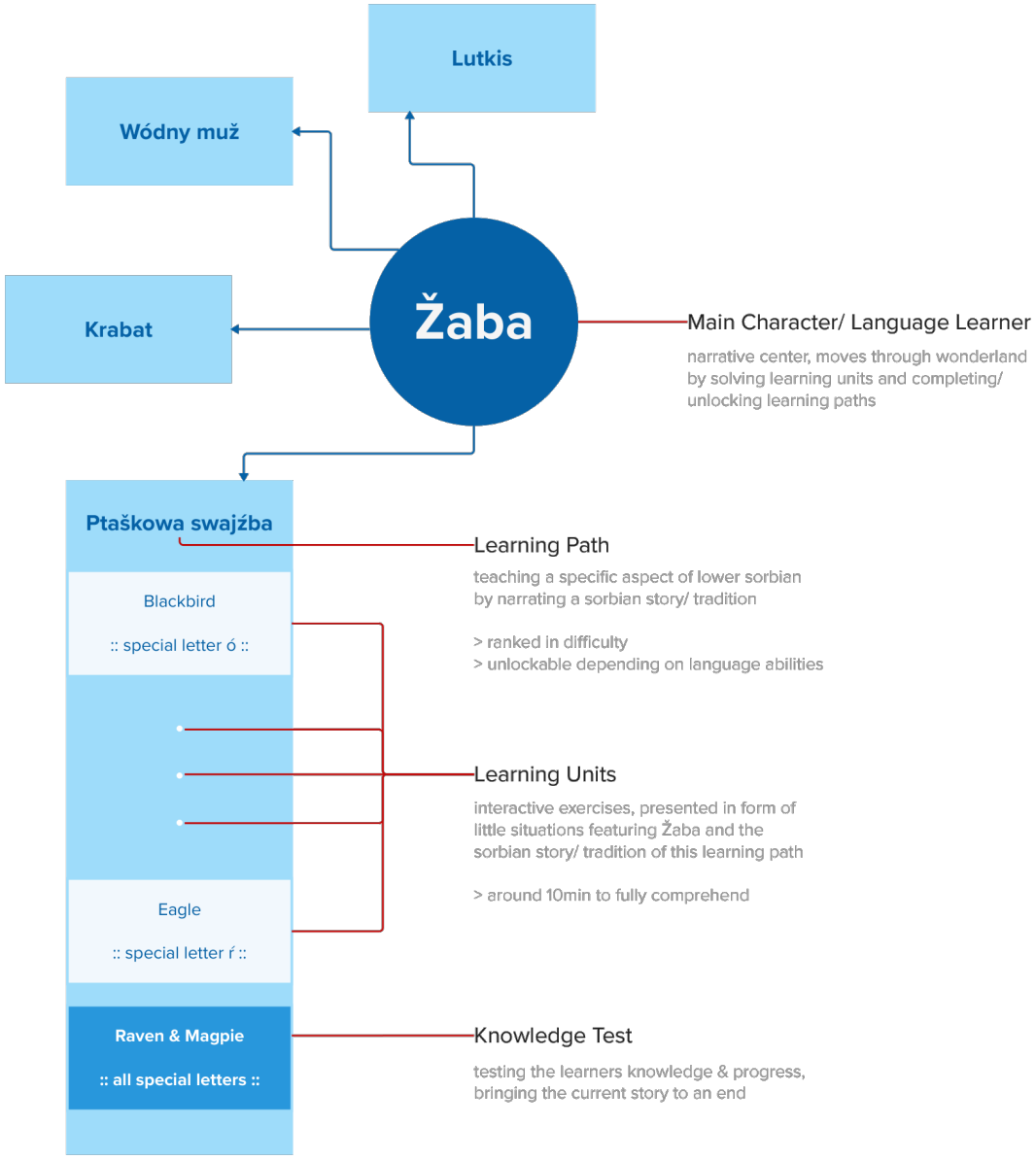
[19] Mozilla Common Voice. https://commonvoice.mozilla.org/de/datasets. accessed: 2024-03-19.

APPENDIX

A. Problem Solution Mind Map



B. Narrative Structure

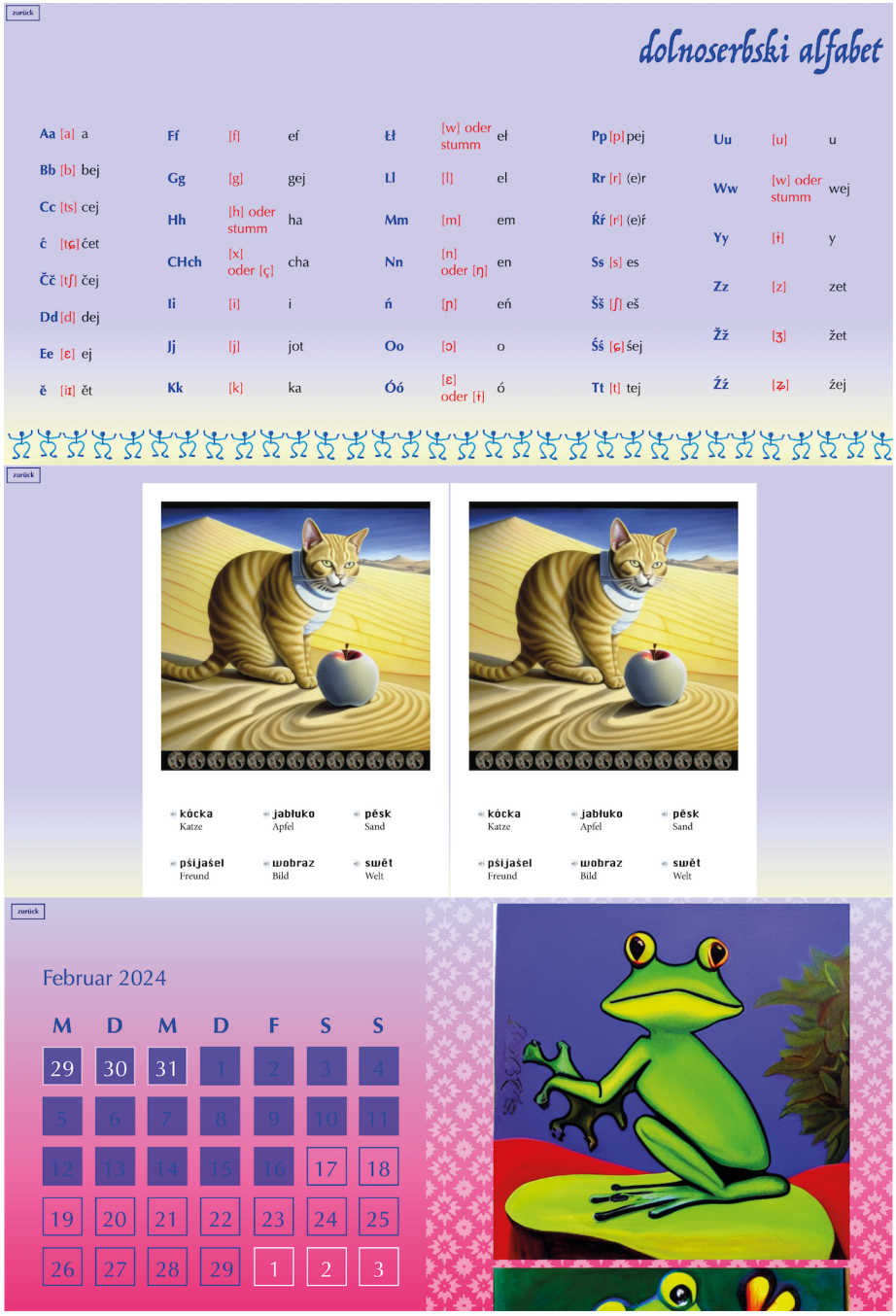


C. Story Visualization

System Screenshots: learning unit 1

D. Main Character Images

E. System Screenshots



lower sorbian alphabet,
weekly vocabulary challenge: ODD PIC overview,
monthly streak view with main character image selection

F. ODD PIC

kócka
Katze

jabłuko
Apfel

pěsk
Sand

pśijaśel
Freund

wobraz
Bild

swět
Welt

G. Identifying Vocabulary Basis

	Corpus	Years covered	Number of tokens
1	Neuestes niedersorbisches Schrifttum	2019 - 2022	2mio
2	Niedersorbisches Referenzkorpus	1990 - 2015	1mio
3	Historisches Schrifttum	current lower sorbian digital library	15mio

Table 1: *Corpus Information*

132_translated

sorbian word	translation	part of speech
KÓŃC	Ende ○ Schluss • Spitze • *až do ≈a* — bis zuletzt • *mimo ≈a* — endlos • *ze wšych ≈ow* — von allen Seiten • *to jo na ≈u to same* — das ist letztendlich das gleiche • *z tym jo něnto ≈* — damit ist es jetzt vorbei • *to změjo zły ≈* — das wird Folgen nach sich ziehen • *z mójim měrom jo było ≈* — um meine Ruhe war es geschehen • *wón njewě, gdy ≈* — er kennt kein Maß • *(krotki) ≈ (wu)gótowaś* — einen kurzen Prozess machen • *na ≈ duce lěto* — das ausklingende Jahr	N
TY	du	P–Pron
NĚKOTARY	1. irgendeiner • mancher ○ *≈ raz* — manches Mal 2. Pl *někotare*: einige • mehrere • manche ○ *~e lěta* — einige Jahre	Ind–Pron
PŚED	1. vor, in Richtung auf etw. ○ *stup se ≈e mnjo!* — tritt vor mich! ○ *až ≈ wrota zajěś* — bis an das Tor fahren 2. • lokal: vor ▪ *≈e mnu stojaś* — vor mir stehen ▪ *≈ wrotami zastaś* — vor dem Tor halten • Zeit: vor ▪ *≈e mšu se pśewoblac* — vor dem Gottesdienst sich umziehen ▪ *≈ spanim cytaś* — vor dem Schlafen lesen • vor, Beziehung zu Personen, Dingen usw. ▪ *≈ rědownju se zagroniś* — sich vor der Klasse entschuldigen	P–Pron
MĚSTO	1. Stadt 2. umg: Platz • Stelle • Stätte • Ort 3. Lage ○ Stelle • Reihenfolge ○ *na prědnem ~će* — an erster Stelle • in erster Linie	N
SŁYŠAŚ	hören • akustisch vernehmen • *wón mało ~y* — er hört schwer • *take něco se wšuźi ~y* — so etw. vernimmt man überall	V
PŚIŚ	1. kommen ○ *(daś) ~žo, kaž ~žo* — mag's kommen, wie es will ○ *co pón ~žo?* — was folgt darauf? ○ *to ~žo ako prědne* — das geht vor ○ *to ~žo jomu drogo* — das wird ihm teuer zu stehen kommen ○ *wón nje~žo z nim sobu* — er kann mit ihm nicht mithalten ○ *masa burow ~že do wótwisnosći* — die Masse der Bauern geriet in Abhängigkeit ○ *to ~žo na jadno* — das läuft auf eins hinaus ○ *stare śerpjenje zasej na nju ~že* — das alte Leiden stellte sich wieder bei ihr ein ○ *to ~žo jomu wjelgin śěžko* — das fällt ihm sehr schwer ○ *ślědkoju to na druge luźe ~žo* — schließlich fällt das anderen zu ○ *glědaj, aby wót kłěba nje~šeł* — pass auf, dass du nicht Arbeit und Brot verlierst! ○ *nje~du na nic lěpšego* — mir fällt nichts Besseres ein ○ *ja njamógu na to ≈* — ich kann mich nicht besinnen ○ *do nimskego ≈* — ins Deutsche fallen ○ *naraz jo ~šło jomu do mysli* — plötzlich ist ihm eingefallen 2. *pó* [4]: etw., jmdn. holen • etw., jmdn. abholen ○ *~žco pó mnjo!* — kommt mich abholen! ○ *~žo pó měchy* — er kommt die Säcke holen 3. umg: werden ○ *brunica dej hyšći droša ≈* — die Braunkohle soll noch teurer werden	V
WE	1. [6] • lokal: in • im ▪ *w njebjach* — im Himmel ▪ *we jśpě* — im Zimmer	Präp

Figure 4: *Result: Frequency Analysis*