Developing Zila: A Spoken Language Resource for the Endangered Slovenian Gail Valley Dialect

Andrej Žgank[1], Gregor Donaj[1], Urh Kolarič[1], Uši Sereinig[2], Tatjana Koren-Zwitter[3], Sanja Boto[3], Sabina Zwitter-Grilc[4], Jasna Vidinić[1], Darinka Verdonik[1]

[1]University of Maribor, Faculty of Electrical Engineering and Computer Science, Maribor, Slovenia
[2]Slovenian Ethnographic Institute Urban Jarnik, Klagenfurt, Austria
[3]Mohorjeva Hermagoras, Klagenfurt, Austria
[4]ORF Kärnten, Klagenfurt, Austria
andrej.zgank@um.si

Abstract

Slovenian is a less-resourced South Slavic language. Existing Slovenian spoken language resources mainly cover the standard language in everyday communication. However, Slovenian encompasses a wide range of dialects, most of which are not represented in available spoken language resources. This paper presents the development of Zila, a Slovenian spoken language resource for the Gail Valley dialect. This dialect is one of the most endangered varieties of Slovenian and is spoken in the extreme north-western periphery of the Slovenian language area. The goal of the project was to build a language resource comprising 100 hours of speech with manually produced transcriptions. The spoken material was collected from members of the Slovenian minority in Carinthia, Austria, with the local community playing a key role in the data acquisition process. A dedicated set of transcription rules was created to capture the full range of acoustic and linguistic features of the Gail Valley dialect, which differs significantly from standard Slovenian. A preliminary speech recognition experiment was conducted to analyze these differences further. The Zila project demonstrates how spoken language technologies can help to preserve the cultural and linguistic heritage of an endangered dialect

Keywords: spoken language resource, Slovenian, endangered dialect, Gail Valley dialect, speech recognition

1. Introduction

Spoken language technologies are becoming integrated increasingly into everyday life, extending their applications from traditional ICT scenarios, such as spoken user interfaces, to advanced IoT-based e-health solutions within intelligent environments (Bhangale and Mohanaprasad, 2021). The growing number of users is reflected in the expanding range of supported languages. Consequently, the development of spoken language resources plays a key role in advancing speech technology, particularly for less-resourced languages.

The Slovenian language is among these less-resourced languages. Although a moderate amount of standard Slovenian spoken language datasets is available for research and development, support for more complex spoken situations remains limited. One particularly challenging aspect concerns dialect. Slovenian is characterized by a high degree of dialect diversity, encompassing 51 dialects and subdialects (Škofic et al., 2011). This diversity stems primarily from the significant role of local communities in the past, historically limited travel opportunities, and the influence of neighboring languages. In some extreme cases, speakers of different Slovenian dialects may even experience difficulty in understanding each other.

Modern societal development is endangering the Slovenian dialect diversity. The increased mobility, the influence of digital technologies and internet content, and the increased role of the urban environment, have had a significant impact on Slovenian dialects. These changes typically reduce the number of active speakers, and limit the opportunities for communication in the dialect. Such trends highlight the importance of documenting and preserving the Slovenian dialects and of incorporating them into future spoken language technologies.

The Slovenian Gail Valley dialect is one of the most endangered among all the Slovenian dialects. It belongs to the Carinthian Slovenian dialect group, which is spoken by the Slovenian minority in the southern parts of Carinthia (Austria) and in the north-eastern parts of Italy (Val Canale), as well as in north-western Slovenia. The Gail Valley dialect represents the extreme northwestern periphery of the Slovenian language area. It is spoken primarily in the municipalities of Feistritz an der Gail / Bistrica na Zilji, Hohenthurn / Straja vas, and the formerly independent municipalites of Egg / Brdo and Vorderberg / Blače in Austria. The Slovenian-speaking minority in these Austrian municipalities constitutes between 5% and 10% of the local population (Austrian Census 2001 statistics, 2025). Their number has declined drastically over the past century, resulting in the dialect's current endangered status.

The primary objective of the Zila project is to preserve the linguistic and cultural heritage of the Gail Valley dialect, and to provide the necessary resources for the development of spoken language technologies. The first part of the project is devoted to constructing a 100-hour manually

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 3325–3332
11–16 May 2026.

transcribed Slovenian Gail Valley speech database, which will be freely available through the CLARIN.SI repository. The second part focuses on developing an automatic speech recognition system for this dialect, which can later be used in various language technology services and applications (e.g. dialect learning, translation to standard Slovenian,…).

This paper presents the process of designing the Zila spoken language resource for the Slovenian Gail Valley dialect, and describes the procedures used for data acquisition and transcription. In addition, it reports on the first speech recognition experiments conducted using a standard Slovenian speech recognizer to assess preliminarily the differences between the complex Gail Valley dialect and standard Slovenian.

The work represents a significant advancement for the Slovenian research community, while also offering potential benefits for work on other endangered dialects or languages. The key contributions are as follow:

- A new 100-hour manually transcribed speech database for the endangered Slovenian Gail Valley dialect.
- A new set of transcription rules designed to handle the challenges of a complex dialect that differs significantly from the standard language effectively.
- A positive example of how community engagement and collaboration with a local cultural organization can boost the data acquisition process, particularly given that the number of potential Gail Valley speakers in Austrian Carinthia with well-preserved dialect is estimated to be fewer than 1,000.

The paper is organized as follows. The existing Slovenian spoken language resources are presented in Section 2. The Slovenian Gail Valley dialect and its characteristics are introduced in Section 3. The data acquisition and manual transcription processes are described in Section 4, while the Zila spoken language resource is introduced in Section 5. The preliminary dialect speech recognition experiments are presented in Section 6, and the conclusions are given in Section 7.

2. Overview of Slovenian Spoken Language Resources

The Slovenian language belongs to the South Slavic branch of the Slavic language family. The systematic development of spoken language resources for Slovenian began in the early 1990s, placing it among the early adopters of spoken language research. Despite these pioneering efforts, however, the quantity and diversity of available Slovenian resources remain limited, typically lagging behind major languages such as English, Mandarin, and Spanish by an order of magnitude. This disparity is due primarily to the relatively small speaker base. There are approximately 2.5 million speakers in total, of whom around 2 million reside in Slovenia. The end effect results in reduced economic motivation for the large-scale development of spoken language technologies.

The evolution of Slovenian spoken language resources can be divided broadly into three periods. The first period, extending from 1990 to 2003, was characterized by efforts to collect isolated and connected speech. This focus reflected the acoustic modeling limitations of early speech recognizers, which necessitated the creation of separate datasets for telephone and studio recordings. Notable resources from this period include FDB 1000 Slovenian SpeechDat(II) (Kaiser and Kačič, 1997; Kaiser et al., 2014), Polidat (Žgank et al., 2002), Gopolis (Dobrišek et al., 1998), VNTV/VNRAD (Žibert et al., 2003), and SNABI database. The total amount of recordings collected during this stage ranged from approximately 10 to 20 hours per resource, accompanied by manually produced annotations and transcriptions.

The second period, covering the years 2004 to 2018, marked a shift toward media and public speech. Representative resources include UMB BNSI Broadcast News (Žgank et al., 2004; Žgank et al., 2014), SiBN Broadcast News (Žibert and Mihelič, 2004), IETK-TV, and the GOS public speech subcorpus (Verdonik et al., 2013; Verdonik et al., 2023c), all focused on broadcast material. In addition, SI TEDx-UM (Žgank et al., 2016) and GOS-Videolectures (Verdonik, 2018; Verdonik et al., 2021) captured public talks and academic lectures. The SloParl corpus (Žgank et al., 2006) contains parliamentary debates held between 2000 and 2005, while the SOFES database (Dobrišek et al., 2017) includes speech queries concerning flight information. The volume of speech data per resource increased to between 30 and 100 hours, though most annotations and transcriptions were still produced manually. Only SI TEDx-UM includes partially automated transcriptions. The relatively modest scale of these resources has constrained the development of end-to-end automatic speech recognition (ASR) systems for Slovenian. Furthermore, their availability remains inconsistent: while some datasets are freely accessible, others are distributed via ELRA/ELDA or remain proprietary.

The third period began in 2019, with the launch of an initiative to develop a freely available 1,000-hour spoken language resource for Slovenian within the framework of the RSDO project (Verdonik et al., 2025). This collection encompasses a diverse range of speech types, including read speech, media recordings, private conversations, and parliamentary debates. The Artur database, available through the CLARIN.SI

repository (Verdonik et al., 2023a; Verdonik et al., 2023b), has been applied successfully in deep learning–based models, achieving results comparable to those obtained for English ASR systems of similar complexity (Slo-Bench ASR, 2025). Although the Artur database represents a major step forward in the development of Slovenian ASR, its size remains approximately an order of magnitude smaller than comparable resources for major languages, which, typically, contain between 10,000 and 100,000 hours of speech.

Dialect speech presents a persistent challenge for automatic speech recognition, due to its distinctive acoustic-phonetic and linguistic characteristics. The existing Slovenian spoken language resources provide only limited coverage of dialect variation. While some non-standard and spontaneous speech samples have been collected, these represent predominantly standard Slovenian. Consequently, there is currently no spoken language resource for Slovenian that includes a substantial quantity of dialect speech.

3. Characteristics of the Slovenian Gail Valley Dialect

The Slovenian Gail Valley dialect is one of the most archaic Slovenian dialects (see Table 1). With some differences, the dialect is spoken along the Gail River from Hermagor (west), while in the east, the Gail Valley characteristics merge into the Rosen Valley dialect in the villages around Lake Faak.

Speakers for the Zila spoken language resource were recruited from the area surrounding Feistritz an der Gail / Bistrica na Zilji in the southern part of Carinthia, Austria. This region represents the core area in which the Slovenian Gail Valley dialect is preserved most faithfully (see Figure 1).

Figure 1: Gail Valley region, where the speakers where recruited – the area marked with light green. Map source: OpenStreetMap, CCBY-SA.

Characteristics from the Old (Common) Slavonic (6th to 8th centuries)

Reflexes of nasal vowels: [en, on] srenča – happiness, obronč – hoop, Standard Slovenian: sreča, obroč

consonant cluster: [dl, tl] vidle, Standard Slovenian: vile – fork

The prefix *vy is realised as [bə] or [bi]: bəlengu sən se (i was born), bivaž (spring)

Vocalism

Long or short vowels [a, e, i, o, u], open or closed vowels [e, ɛ, o, ɔ] + mid central vowel [ə]

Pronounced diphthongization: e > iə, e > eə, i > iə, o > uə, o > oə, o > uo

»Akanje«: unaccented or briefly accented [o] becomes [a]: pod goro > pad garo, dolina > dalina

Consonants, sonants

Due to the many voice changes and palatalizations, the Slovenian Gail Valley dialect is difficult to understand.

l > v ("švapanje"): [l] becomes bilabial [v] before [a, o, u]: šla > šva (went)

k > č, before the front vowels [e, i]: kikla > čikla (skirt, underskirt), kje > čej (where)

k > c, in the locative before the ending [e]: v potoce (in the brook)

h > š, before [i, u]: hiša > šiša (house)

g > j/ž, before [e/ə]: noge > nuəjə (feet), šiling > šilinžə (shillings)

m > n, at the end of a word: vem > viən (i know), imam > man (I have)

v > b, the labiodental [v] is often pronounced almost like the plosive [b]: krave > krabe (cows)

[j] and [v] are often lost at the beginning of words: jesti > estə (to eat), volja > ola (the will)

	Transcription
Gail Valley	V tistah časah sa sə večbart na goərah, pər Zilə bəl pər patokə pa na Ogə čudnə rečə zgadilə.
Standard Slovenian	V tistih časih so se večkrat na gorah, pri Zilji, ob potokih in na Logu godile čudne reči.
English	In those days, strange things often happened in the mountains, at Zilja, by the streams and at Log.

Table 1: An example of Slovenian Gail Valley dialect text with translations into standard Slovenian and English from the Mijalca majalca fable (Bartoloth M., 2021).

Vocabulary

The basic vocabulary contains many words that differ completely from the Slovenian language or other Slovenian dialects: žabaritə (to speak), ščəkatə (to milk), buratə (drive), pruno (blue), črnjevə (red), rusə (brown), ženč (bad, rotten).

Due to centuries of linguistic contact with the German-speaking world, the Slovenian Gail Valley dialect also contains many German loanwords, some of which date back to the 10th to 13th centuries. Many words also come from Italian, and they aren't usually perceived as foreign words.

4. Speech Acquisition and Transcription

The number of speakers of the Slovenian Gail Valley dialect is very limited, which necessitated the exploration of all possible sources for speech data acquisition. Recordings were first collected from regional and national media organizations, ethnographic archives, and media publishers (i.e. ORF Kärnten, ORF, ZRC SAZU, SNI Urban Jarnik, Podlipnik film, SPD Zila). In the second phase, field recordings were conducted within the Gail Valley communities, where the speakers were recorded in their private environments. The recruitment of potential speakers was supported strongly by Slovensko prosvetno društvo Zila (SPD Zila, eng.: Slovenian cultural association Zila) a local Slovenian minority cultural organization that is dedicated to preserving the community's cultural and linguistic heritage. A variety of audio technologies were employed, due to the diverse origins of the speech recordings. The oldest recordings, dating back more than 50 years, were sourced from analog magnetic tapes and video cassettes, while CDs and DVDs represented the digital media. In the case of old media, two challenges arose. First, we had to find a suitable device, and second, the audio quality (i.e., signal-to-noise ratio, background noise, frequency characteristics, etc.) was usually lower than in recent recordings. Most of the field recordings were made using digital voice recorders and smartphones. The characteristics of each recording session (e.g., the recording device and environment) were documented and used subsequently for a quality analysis.

An FTP server was established to collect the speech recordings, speakers' consent forms, and transcriptions. The users were assigned different roles: recorders, transcribers, and admins. Their roles gave them different privileges for viewing and adding data to the server. The server was set up to prevent the users from deleting any files accidentally. When the files were trying to be overwritten, the system renamed them automatically, and thus preserved both versions. Only the server administrator was able to delete files.

The uploaded audio files were checked for their basic audio parameters. The preferred format for upload was 16-bit signed integer PCM, with one channel (mono) and a sample rate of 16 kHz.

After upload, the files were assessed manually by listening to them to check their overall sound quality and speech clarity. Some files were discarded due to a low sound level or background noise. The approved files were copied to another folder, after which the transcription process began.

The transcription of the Zila dialect corpus followed a single-level extended orthographic scheme and was done with the Transcriber 1.5.1 (Barras et al., 2001), which ensures precise alignment between the transcript and the corresponding audio signal. The speech was segmented into basic transcription units (segments) at pauses of approximately 0.2 seconds or longer, unless this would result in unnaturally short or excessively long segments; longer pauses over 1.5 seconds and extended unintelligible speech were marked as separate segments. Each segment was assigned to its originating speaker with a unique speaker ID, and overlapping speech was indicated through specific segment markings distinguishing the first and second speakers.

The transcription preserves all the morphological and syntactic features of the spoken dialect and records speech verbatim, including hesitation phenomena, fillers, false starts, and other disfluencies. The words were transcribed in their dialectal forms, with consistent orthographic representation of the characteristic phonological features (see Table 1). The punctuation followed standard orthographic rules, but excluded punctuation marks not supported in Transcriber.

Non-verbal and paralinguistic sounds—such as laughter, coughing, or background noise—were inserted as events. Unintelligible were annotated as unintelligible and incomplete words with empty parentheses for truncated fragments. Sensitive personal data were enclosed in square brackets and marked temporally in the audio section to enable later anonymization.

The transcription was performed by two linguists participating in the project. Prior to the submission of each transcription, the transcribers conferred to resolve any issues encountered during the process, particularly those related to the representation of dialectal features. In the final stage, all the transcriptions underwent an additional review, to verify their technical compliance with the established transcription guidelines.

5. Zila Spoken Language Resource

The Zila spoken language resource comprises 100.5 hours of speech recordings accompanied

by manually generated transcriptions. The demographic characteristics of the speakers are summarized in Table 2.

The focus on a narrowly defined geographic area—where the Gail Valley dialect is preserved best—resulted in a relatively small number of speakers. In total, the dataset includes 75 speakers, of whom 65 spoke the Gail Valley dialect. The remaining ten speakers used either standard Slovenian or other Slovenian dialects and acted primarily as interviewers, with smaller amounts of speech. The age distribution was partially outside the optimal range for deployment in speech technology applications. Specifically, 30 speakers were elderly (over 65 years), a group in which voice changes related to aging can be expected. The oldest speaker was 95 years old.

Total number of speakers:	75
Number of Gail Valley dialect speakers:	65
Gender:	F: 35, M: 30
Age:	24 -- 95
Accent preservation (scale 1—3)	2.91

Table 2: The demographic characteristics of speakers in the Zila spoken language resource.

This compromise was justified by the fact that the elderly members of the community tend to preserve their dialect more faithfully, having been less exposed to the influence of other languages and mass media, which is far more prevalent among younger speakers.

The gender distribution of participants was 35 female and 30 male speakers. The vast majority of the speakers were members of the Slovenian Gail Valley minority, residing in the area around Feistritz an der Gail / Bistrica na Zilji. The degree of accent preservation was assessed subjectively by the transcribers during the transcription process using a three-point scale. The average accent preservation score was 2.91. This outcome reflects the carefully designed data collection process, in which speakers with well-preserved accents were sought actively within the community. Except for one individual, all the speakers self-identified the Slovenian Gail Valley dialect as their native language. All the participants were bilingual, with Austrian German as their second language.

Of the first 63 hours of speech recordings, 41 hours across 123 recording sessions have been selected. These data were used for analysis and preliminary evaluation of automatic speech recognition. The manual segmentation produced 28,922 audio files, with the longest recording lasting 22.84 seconds. The split between the training and test sets was 90% to 10%. The corresponding transcriptions contained 297,000 words, of which 27,200 are distinct. The relatively large vocabulary size reflects the morphological richness of the Slovenian language, which is particularly evident in dialectal speech.

Set	**Train**	**Test**
Duration	4.64 s	4.67 s
RMS	0.056	0.056
SNR	31.04 dB	30.82 dB
Clipping ratio	0.017	0.015

Table 3: Average audio characteristics (duration, RMS, SNR, and clipping) across the recordings in the training and test sets.

To verify the suitability of the dataset and ensure a balanced distribution between the training and test subsets, several basic acoustic metrics were calculated for all the recordings. Table 3 presents the average values of duration, root mean square (RMS) amplitude, signal-to-noise ratio (SNR), and clipping ratio for both subsets. The mean duration of the recordings was approximately 4.6 s, with nearly identical values in the training and test sets, indicating consistent segment lengths. The RMS and SNR values were also comparable, suggesting a similar overall signal level and noise characteristics across the subsets. Finally, the low clipping ratio (<0.02) confirms that the recordings were not affected by significant signal saturation. These results demonstrate that both subsets are acoustically well balanced and suitable for automatic speech recognition.

6. Preliminary Automatic Speech Recognition Evaluation

The current version of the Zila spoken language resource was employed for a preliminary evaluation using a general Slovenian automatic speech recognition (ASR) system. The objective of this evaluation was to examine the impact of heavily accented speech on recognition accuracy for Slovenian — a linguistically complex and morphologically rich language characterized by a high degree of inflection and relatively free word order.

The Zila resource was divided provisionally into training and testing subsets. However, only the test subset was used in the present experiments, as no model training or fine-tuning on the Gail Valley dialect was performed at the time. The test subset represented 10% of the full dataset, corresponding to 2,893 audio recordings. The files were selected randomly without isolating

speakers or recording sessions explicitly within a single subset.

The Slovenian ASR system used for evaluation was trained on the Artur speech database (Verdonik et al., 2025), a freely available 1,000-hour resource that can be obtained from the CLARIN.SI repository (Verdonik et al., 2023a; Verdonik et al., 2023b). The Artur corpus was developed between 2019 and 2022, and includes recordings and manually produced transcriptions across four speech domains: public and media speech, parliamentary debates, non-public conversational speech, and prepared read speech.

The system was implemented using the WeNet framework (Yao et al., 2021), applying a transformer-based ASR model for Slovenian. The model architecture consisted of a conformer encoder with 12 blocks, each containing 4 attention heads and an output dimension of 256. A transformer decoder was applied comprising 6 blocks with 4 attention heads. The model training utilized the Adam optimization algorithm.

The evaluation performance was assessed using two standard metrics: word error rate (WER) and character error rate (CER). The general Slovenian ASR system was evaluated first on the Artur test set as the baseline (see Table 4).

Set	WER(%)	CER(%)
Artur test set	5.75	1.87
Zila dialect speech	80.78	42.10

Table 4: The Slovenian speech recognition results with the Artur and Zila test sets.

The transformer-based end-to-end speech recognition system achieved a word error rate (WER) of 5.75% on the Artur test set, which is comparable to other speech recognition systems of similar complexity designed for general speech recognition. The corresponding character error rate (CER) for the Artur test set was 1.87%, consistent with the performance typically achieved by well-trained general-purpose ASR systems.

In contrast, testing with the Zila dialect speech dataset resulted in a substantial degradation of performance, with the WER increasing to 80.78%. This WER degradation was caused by the differences between standard Slovenian and the heavily accented Gail Valley dialect. The discrepancy between the two test sets was less noticeable in the case of CER, where the Zila test set achieved 42.10%.

A detailed analysis of the recognition errors showed that the ASR system, trained exclusively on standard Slovenian speech, failed to model the specific phonetic characteristics of the dialect speech accurately. Frequently, the accented phoneme is mapped to a similar general phoneme from a broader acoustic category. Furthermore, the analysis indicated that the differences in word boundaries between standard Slovenian and the dialect variant also contributed to the high level of recognition errors.

The preliminary speech recognition evaluation served a twofold purpose: first, to assess the usability of applying an existing standard Slovenian speech recognizer, and second, to verify that the proposed data acquisition procedure, which originated primarily within the limited private sphere, can result in a dataset of sufficient quality for speech technology research. The achieved character error rate supported the validity of the data acquisition procedure.

At the same time, the analysis demonstrated clearly that standard Slovenian speech recognition systems cannot be applied directly to tasks involving such a complex dialect, where an additional set of transcription conventions, beyond standard Slovenian, was defined. It is clear that dialectal ASR models are needed to address this challenge.

Consequently, future research in the field of automatic speech recognition will focus on improving the acoustic modeling of distinctive acoustic-phonetic characteristics through fine-tuning or modeling of pronounced variants. The observed discrepancies between the WER and CER values, along with the detected word boundary errors, suggest that applying language modeling methods could improve speech recognition performance. However, it remains to be determined whether a sufficiently large body of Slovenian Gail Valley dialect text material exists (besides the available audio transcriptions) to support the development of a usable text corpus. Also, the possibility of evaluating speech recognition results at the combined phrase and character levels will be considered.

7. Conclusion

The development of the Zila spoken language resource represents an important step toward documenting and preserving the Slovenian Gail Valley dialect, one of the most endangered and least resourced varieties of Slovenian. Building upon three decades of spoken language resource development for Slovenian, the project introduces several key innovations: a newly collected corpus of private and spontaneous speech, a carefully designed transcription protocol tailored to the phonological and morphological characteristics of the dialect, and a workflow that integrates the expertise of trained linguists with the active

involvement of local cultural organizations. This collaborative model demonstrates how community-based initiatives can contribute substantially to the creation of linguistic resources for endangered dialects.

The Zila project extends the existing Slovenian transcription conventions through a detailed orthographic system that captures dialectal features accurately while remaining compatible with other spoken language corpora. The verified transcriptions form a reliable basis for further annotation and computational analysis. The preliminary ASR experiments using a general Slovenian model revealed the complexity of recognizing heavily accented and dialectal speech, emphasizing the need for dialect-specific data and adapted modeling approaches.

The corpus will be made freely available through the repository of the CLARIN.SI consortium[1]. Future work will focus on developing specialized ASR models to accommodate dialectal variation better, thereby advancing both technological inclusion and the preservation of linguistic heritage.

8. Acknowledgments

This work is part of the LINGUA Project, funded by the EU Interreg Program Slovenia-Austria 2021–2027. In addition to the lead partner, the Hermagoras Association in Klagenfurt, Gorenjski Glas from Kranj and the Škofja Loka Adult Education Centre are participating as project partners. This research work was funded partially by the “Large Language Models for Digital Humanities” (GC-0002) and ”Basic Research for the Development of Spoken Language Resources and Speech Technologies for the Slovenian Language” (J7-4642) research projects and partially by the “Advanced Methods of Interaction in Telecommunication” (P2-0069) research program by the Slovenian Research and Innovation Agency (ARIS). The authors wish to thank the members of the Slovenian minority in the Gail Valley, Austria, who contributed their speech to the Zila spoken language resource.

9. Bibliographical References

Austrian Census 2001 statistics (2025). Available: https://www.statistik.at/en/statistics/population-and-society/population/population-stock/historic-censuses

Barras, C., Geoffrois, E., Wu, Z., & Liberman, M. (2001). Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication*, 33(1-2), 5-22.

Bartoloth M. (2021). *Mijalca majalca: Zilska basən / Ziljska basen / Ein Gailtaler Märchen*. Klagenfurt, Vienna and Ljubljana: Mohorjeva Hermagoras.

Bhangale, K. B., & Mohanaprasad, K. (2021). A review on speech processing using machine learning paradigm. *International Journal of Speech Technology*, 24(2), 367-388.

Dobrišek, S., Gros, J., Mihelič, F. and Pavešić, N. (1998). Recording and labelling of the GOPOLIS Slovenian speech database. *First International Conference on language resources & evaluation*: Granada, Spain, 28-30 May 1998 (1089—1096). European Language Resources Association.

Dobrišek, S., Žganec Gros, J., Žibert, J., Mihelič, F. and Pavešić, N. (2017). Speech Database of Spoken Flight Information Enquiries SOFES 1.0.

Kaiser, J. and Kačič, Z. (1997). SpeechDat (II) Slovenian Database for the Fixed Telephone Network. Maribor, Slovenia: University of Maribor.

Slo-Bench Automatic Speech Recognition leaderboard (2025). Available: https://slobench.cjvt.si/leaderboard/view/10.

Škofic, J. et al., (2011). *Slovenski lingvistični atlas*. 1, Človek (telo, bolezni, družina): 2. Komentarji (Vol. 1). Založba ZRC.

Verdonik, D., Kosem, I., Vitez, A., Krek, S. and Stabej, M. (2013). Compilation, transcription and usage of a reference speech corpus: The case of the Slovene corpus GOS. *Language resources and evaluation*, 47(4), 1031—1048.

Verdonik., D. (2018). Korpus in baza Gos Videolectures. In *Zbornik 11. konference Jezikovne tehnologije*.

Verdonik, D., Bizjak, A., Žgank, A., Maučec, M. S., Trojar, M., Gros, J. Ž., ... & Dobrišek, S. (2025). Strategies for managing time and costs in speech corpus creation: insights from the Slovenian ARTUR corpus. *Language Resources and Evaluation*, 59(3), 1899-1924.

Yao, Z., Di Wu, Wang, X., Zhang, B., Yu, F., Yang, C., ... & Lei, X. (2021). WeNet: Production Oriented Streaming and Non-Streaming End-to-End Speech Recognition Toolkit. In *Interspeech* (Vol. 2021, pp. 4054-4058).

Žgank, A., Kačič, Z. and Horvat, B. (2002). Preliminary evaluation of Slovenian mobile database PoliDat. *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC’02)*.

Žgank, A., Rotovnik, T., Sepesy Maučec, M., Verdonik, D., Kitak, J., Vlaj, D., Hozjan, V., Kačič, Z. and Horvat, B. (2004). Acquisition and annotation of Slovenian broadcast news database. *Fourth international conference on language resources and evaluation, LREC 2004*, Lisbon, Portugal.

Žgank, A., Rotovnik, T., Grašič, M., Kos, M., Vlaj, D., and Kačič, Z. (2006). Sloparl-Slovenian parliamentary speech and text corpus for large

[1]Slovenian CLARIN.SI repository: https://www.clarin.si/repository/xmlui/

vocabulary continuous speech recognition. *Ninth International Conference on Spoken Language Processing*.

Žgank, A., Sepesy Maučec, M. and Verdonik, D. (2016). The SI TEDx-UM speech database: A new Slovenian spoken language resource. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (4670—4673).

Žibert, J., Martinčić-Ipšić, S., Ipšić, I. and Mihelič, F. (2003). Bilingual speech recognition of Slovenian and Croatian weather forecasts. *Proceedings EC-VIP-MC 2003. 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications* (IEEE Cat. No. 03EX667) (Vol. 2, pp. 637—642). IEEE.

Žibert, J. and Mihelič, F. (2004). Development of Slovenian broadcast news speech database. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*.

10. Language Resource References

Kaiser, Janez; et al., 2014, Slovenian SpeechDat(II) FDB-1000, ISLRN: 009-566-242-903-2, http://catalog.elra.info/product_info.php?products_id=188

Verdonik, Darinka; et al., 2021, Spoken corpus Gos VideoLectures 4.2 (transcription), Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1444.

Verdonik, Darinka; et al., 2023a, ASR database ARTUR 1.0 (audio), Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1776.

Verdonik, Darinka; et al., 2023b, ASR database ARTUR 1.0 (transcriptions), Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1772.

Verdonik, Darinka; et al., 2023c, Spoken corpus Gos 2.1 (transcriptions), Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, http://hdl.handle.net/11356/1863.

Žgank, Andrej; et al., 2014, Slovenian BNSI Broadcast News Speech Corpus, ISLRN: 502-280-144-938-4, http://catalog.elra.info/product_info.php?products_id=1064