A 75-year-old Hungarian spontaneous speech database

M. Gősy

Research Institute for Linguistics, Hungarian Academy of Sciences
gosy.maria@nytud.mta.hu

Abstract: The first attempt to develop a large collection of recorded speech material in Hungarian was made by the phonetician Lajos Hegedűs in the 1940s. He wanted to preserve the sounding of the various Hungarian dialects in the country and even outside Hungary with the purpose of analyzing, among other things, the intonation, pauses and rhythm of speech in those dialects. This paper introduces the properties of the Hegedűs Archives on the one hand, and discusses the results of investigations that aimed to compare the speaking strategies of people that could be the great-grandparents of the speakers living now, on the other hand.

1. Introduction

The proverb What is spoken flies, what is written never dies is known in several languages. However, the possibility of sound recording from 1888 onwards may question the truth of the human experience summarized in that proverb. A relatively short time – about 100 years – elapsed between Edison’s invention and the development of the technical background that now provides a possibility to record and save huge speech materials for phonetic and speech technology research.

Language and language use continually change due to the fact that languages live as long as the communities that speak them remain in existence. The quality and temporal course of changes also differ across periods since they are determined by social and other human factors [e.g. 1, 2]. The recognition of ongoing changes is no easy task; quite often, it is difficult to tell what is and what is not an instance of language change [3, 4]. The further away we are from a period, however, the more exactly it can be determined what qualifies as a change as compared to the present situation, what tendencies can be identified, and what possible causes may have been instrumental in it. In the case of spoken language, this is possible only if we have access to recorded speech materials that can be analyzed in multiple ways and hence compared to facts of the present state of the language.

Speech materials that were recorded for a particular purpose are immensely important. The development of large databases is due to the evolution of technology: to the emergence of computers with a large memory capacity, and the related demand for large amounts of texts recorded with specific aims in mind and encoded in a uniform manner so that researchers can analyze them in various respects. Speech databases in the current sense go back to a few decades only, both internationally [e.g. 5, 6, 7, 8] and in Hungary [e.g. 9, 10]. However, present-day corpora and databases have forerunners that are worth knowing about. We can learn from the mistakes that early developers committed as well as from their advantages and actual achievements, from their successes and failed attempts, from the knowledge they have amassed.

The first fragmental Hungarian dialect recordings we know of are found in the folk song recordings made by Béla Bartók and Zoltán Kodály, the famous Hungarian composers. In 1912, the Benedictine monk Ányos Biró recorded dialectal speech in various locations in
Hungary specifically with the aim of handing Hungarian dialects down to posterity [11]. Probably the earliest database-like collection of spoken language recordings can be linked to the name of the phonetician József Balassa (1864–1945). The recordings were made during the First World War, under the auspices of the Phonetic Laboratory of the Academy of Eastern Trade led by Balassa. In a detention centre near Esztergom, the speech of Votyak, Bashkir, Tartar and Russian prisoners of war was recorded. The total number of subjects was 48. We know that the recordings were made to a wax cylinder by a portable phonograph. However, their cataloguing and analysis were not feasible during the war, and we do not know anything concerning their whereabouts; probably they were destroyed [12]. Yet the mere fact that such a project was carried out suggests that our ancestors felt the necessity of large-scale recording of speech.

Some three decades later, another Hungarian phonetician Lajos Hegedűs started recording dialect speech: this gave rise to the earliest extant collection in this country of speech recordings involving many speakers, known as the Hegedűs Archives today.

2. The Hegedűs Archives

2.1. The founder

It was in the nineteen-forties that, at the initiation of Lajos Hegedűs, regular production of dialect sound recordings began. Lajos Hegedűs was born on 1 March, 1908, and died young on 23 July, 1958. He was a phonetician and linguist, a disciple of Zoltán Gombocz at Budapest University, and he also studied in Vienna and in London. In 1943, he became an unpaid professor at Debrecen University; from 1944 to 1950 he worked for the Transdanubian Research Institute in Pécs (South-Western Hungary). In 1950 he was employed by the Research Institute for Linguistics of the Hungarian Academy of Sciences, and created “the museum of Hungarian dialects” with his dialect recordings. He was one of the most eminent early representatives of experimental phonetics in Hungary, and his papers are outstanding in an international perspective, too. It cannot be an accident that it was him who insisted on the importance of the creation of a collection of dialect speech recordings.

The aim of the project was to record speech materials at various locations within Hungary and in surrounding countries, in a variety of genres: narratives, storytelling, spells, etc. The aim was not a systematic collection of diverse speech styles: the main point was to record and preserve the various local dialects of Hungarian.

It is hard to believe today that Hegedűs had to insist that the significance of sound recordings, as opposed to data recorded on paper, lies in the fact that they also include suprasegmental features of speech. Let us quote one of his arguments from 60 years ago: “The material collected on gramophone records can be transposed to a kymograph by the help of an electromagnetic apparatus and it can be submitted, in whole or part by part, to objective study concerning duration, intonation, pauses, or rhythm...” [13, p. 6]. Thus, Hegedűs was guided by three aims: to preserve samples of spoken dialects, to provide sound materials for experimental phonetic study, and (especially) to create the possibility of analyzing prosodic features of speech. Figure 1 shows the kymograph.

The first recording was made by Lajos Hegedűs in August 1940 in Drávacsehi (Baranya County). The speaker was András Száva, a 78-year-old man, reciting a tale of the Willowy Princess (8 min 22 sec). Initially, Hegedűs’ work was assisted by the linguist Mihály Temesi (55 recordings), then he made recordings on his own between 1947 and 1954. Later on, other linguists joined in: József Végh, Lajos Lőrincze and Samu Imre.
The latter two also used the questionnaire of a dialect atlas then under preparation, and made sound recordings of the answers of the subjects. 58 recordings were made by the linguists Kálmán Keresztes and Lajos Balogh; in the recording of one subject each, the phonetician Iván Fónagy (later professor at Sorbonne) and the linguist György Szépe also participated. However, the name Hegedűs Archives was not randomly given, decades later, to the collection: Lajos Hegedűs personally made a total of 1538 recordings (92% of the whole material).

2.2. The subjects

The total number of subjects was 686. Due to reasons that remain unknown to us, the age of some of the subjects was not recorded, or else this piece of data did not survive in all cases. Typically, whole families were interviewed simultaneously, in some cases three generations’ speech is found in a single recording. Most subjects were adults, and many were over 40 (Table 1). The oldest speaker was an 88-year-old woman; the youngest was a little girl of 10. The number of subjects under 18 years of age was 16.

Table 1: Subjects’ ages and genders in the Hegedűs Archives.

<table>
<thead>
<tr>
<th>Age range (years)</th>
<th>Number of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>females</td>
</tr>
<tr>
<td>71 – 88</td>
<td>46</td>
</tr>
<tr>
<td>61 – 70</td>
<td>30</td>
</tr>
<tr>
<td>51 – 60</td>
<td>40</td>
</tr>
<tr>
<td>41 – 50</td>
<td>52</td>
</tr>
<tr>
<td>31 – 40</td>
<td>34</td>
</tr>
<tr>
<td>19 – 30</td>
<td>21</td>
</tr>
<tr>
<td>16 – 18</td>
<td>6</td>
</tr>
<tr>
<td>10 – 15</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>241</td>
</tr>
</tbody>
</table>

The number of speakers whose age is known is 414; we do not know the ages of over 150 subjects, and in a number of cases even the name of the speaker is missing, labels like “female voice”, “male voice”, or “unknown” were applied. We also find labels saying “several speakers”: this covers conversations and/or simultaneous communication but the names and
ages of the participants are missing (e.g. women chatting while spinning together). Women figure far more prominently than men do; the number of women/girls whose ages are known is 241, that of men is 173. Talkativeness, it appears, was a characteristic of women; but another reason for this asymmetry may have been that the men were doing jobs that hindered them in participation more extensively. Female subjects normally worked at home or else they worked as day labourers. The men were mainly agricultural workers, but blacksmiths, carpenters, herdsmen, etc. were also found among them.

2.3. The recordings

The total number of recordings is over 1600. They were made in a number of counties within Hungary and in areas outside the country where native Hungarians lived, in Czechoslovakia, Romania, the Soviet Union, or Austria. The fieldwork sites were in the following counties (with the number of recordings made there in parentheses): Baranya (536), Hajdú-Bihar (84), Vas (54), Nógrád (17), Tolna (394), Pest (70), Fejér (7), Zala (176), Veszprém (77), Szabolcs-Szatmár (101), Bács-Kiskun (11), Borsod-Abaúj-Zemplén (1), Maros-Torda (1), Heves (23), Nyitra (38), Abaúj-Torna (36), Gömör and Kishont (7), Ung (6), Krassó-Szörény (2), Kolozs (2), Csík (1), Somogy (3), Csongrád (19), Burgenland (1). The counties are not represented equally, and the numbers of sites in each differ widely. In some counties, 10 or more villages served as collection sites, while in some others only one or two did. The highest number of recordings comes from Baranya County (South-Western Hungary). Very few recordings were made e.g. in Fejér County (roughly the middle of the country, west of the Danube). All recordings made in Csongrád County (South-Eastern Hungary) come from a single village, Tápé. In Somogy County (west of the Danube, situated centrally) recordings were only made in Somogyhatvan, in Csík County (today in Romania), only in Csíkmadaras, and in Pest County (around the capital) only in Páty.

Recordings were dated: usually just the year was given, less frequently the month as well, and in several cases the date was given exactly, including the day when the recording was made. The first recording was made, as already mentioned, in August 1940; the last one was made on 3 May 1957. Grey (later, pink) cards of size A/4 (14.8 × 21 cm) were made to accompany the records, containing many important pieces of information. Along with the name of the collector, these cards also revealed who produced the lacquered records themselves. The Hegedűs Archives include a total of over 50 hours of sound recordings (approximately 600,000 running words). The duration of the shortest recording is 2 minutes, and that of the longest is 14 minutes. The topics are varied. In terms of spontaneity vs. interpretation (of pre-existing texts), they can be classified into three broad categories.

(i) Narratives: the speaker talks about an event related to his/her life, family, or job. These topics are usually raised by the field worker, and the subject talks about them spontaneously, without any previous preparation. For instance: folk customs, superstitions, festive occasions, stories from the life of the village, ways of baking bread or cakes, description of a pig-killing feast, stories on military recruitment, life stories, everyday jobs (fishing, sowing, harvest, shirt making, tin moulding, cheese making, sheep raising, etc.), wedding feasts, saint’s-day fairs, healing, soldiers’ stories, dispelling rats, thefts, death cases, kid’s rigs, maypole raising.

(ii) Texts of varying length handed down from generation to generation in a fixed form like spells, wedding rhymes, Easter toasts, laments, and nursery rhymes. (Occasionally, the recordings also include sung sequences.)

(iii) Semi-spontaneous (or: semi-interpretative) texts where the speaker relates a story that is more or less fixed in terms of both content and linguistic form. For instance: tales, ghost stories, jokes, stories about scamps.
Narratives are sometimes interspersed with dialogues, mainly those between the subject and the field worker, but also real conversations in the sense that members of the family or neighbours who are present at the recording may interrupt the conversation every now and then.

2.4. The technology of recording

Recording involved various tools (microphones, amplifiers, power supply, record cutter, raw gramophone records, headphones, etc.); and the sessions typically required a room-sized space (Fig. 2). In the beginning, glass and aluminium-based, decilith-covered or lacquered records were used (Fig. 3).

![Figure 2: Typical circumstances of speech recording in the mid-twentieth century](image)

Recording was made such that the text spoken into the microphone was directly transmitted to the record-cutting machine, and the record was immediately ready to be played. The jackets of the individual records (Fig. 4) and (paper labels of) the records themselves had the most important data written on them: the name of the field worker, the place and time of recording, the name of the subject, the topic, and the technical data of the record produced. Some jackets exhibit Lajos Hegedűs’ own handwriting. About 1700 separate recordings were made by Hegedűs and his colleagues, and they are contained in 842 original records.

![Figure 3: An original lacquered record and a turntable apparatus able to play it (still in working order at the time of writing)](image)

The quality of recordings varies. Most of them are of good quality, very moderately noisy; they are perfectly fit for instrumental analysis (see Fig. 5). In some cases, the recording is rather noisier, due to various reasons, or contains external noises (like the barking of dogs). Some of the latter are still directly analyzable, but there are others that can only be analyzed acoustically after they are stripped of background noise. But even these noisier recordings are clearly comprehensible, hence perfectly fit for linguistic or folkloristic study.
2.5. Archiving

Most of the records were subsequently stored in the Research Institute for Linguistics of the Hungarian Academy of Sciences. Some others recently emerged as they had been in the personal possession of various linguists. Over the decades, the records started to be mildewed, and the number of record players that could play them has diminished considerably by now. The preservation of the material of the Hegedűs Archives in contemporary data carriers became rather urgent by the beginning of the twenty-first century in more respects than one. The process of archiving consisted in two parts. 1. Technical preservation of the records, meaning both that they are physically restored and that their content is saved in contemporary storage devices in order to be playable. 2. Cataloguing the sound material, and compiling its relevant metadata in an electronically searchable database.

1. A grant from the Ministry of Culture (2003–2006) made it possible to restore and save 842 records. The restoration was made difficult by the fact that special expertise was required for cleaning operations so that neither the records nor the recordings should suffer any harm. First, the records were freed of mildew and any other sediment with a special chemical solution, then they were played by an appropriate record player and their content was recorded in a computer file. After that, the new recordings were copied, in audio format, onto CDs and
other external storage devices. The cross-compliance of the CDs and the original records was taken care of, and a system of labelling was devised for the CDs. The whole archival material is now found on the storage server of the Research Institute for Linguistics of the Hungarian Academy of Sciences, as well as on 122 CDs and two external hard discs.

2. At the same time, the metadata of the recordings were included in computer files on the basis of the notes that accompanied the original records (using the Access program). This searchable information database contains the following: serial number of the CD containing the restored sound material, track number of the CD containing the recording in question, serial number of the original record, date and place of the sound recording session (both the name of the village and that of the county), name of the subject (if available), age of the subject (if available). Furthermore, the database includes the duration of the recording (in seconds), the name of the person who made the recording, the actual topic(s) of the recording, and its definition in terms of a pre-established list of speech topics. The last-mentioned piece of information concerns the exact contents of the given recording. The topic labels are as follows: narrative, part of tale, superstition, conversation, rhymed tag, (nursery) rhyme, toast, soldier’s story, dirge, dialect words (questions and answers of the dialect atlas in preparation at the time of recording), story, recital, song, and poem. These categories make it possible to search the material in terms of speech topics.

3. Studies carried out so far on the material of the Hegedűs Archives

The first portions of the Hegedűs Archives to be transcribed and analyzed by folklorists were some recordings of folk tales. The phonetic study of the recordings started in 2005 and goes on ever since. Linguistic studies concerned disfluency phenomena, prosodic features of spontaneous speech, and a comparison of various speech genres in terms of their phonetic characteristics.

Types and proportions of occurrence of disfluency phenomena were studied in recordings of 28 speakers (14 female and 14 male ones, aged between 25 and 80). The data were compared to instances of disfluency in present-day spontaneous speech by standard speakers of the same ages and genders [14]. In selecting the latter subjects, care was taken that their speech rate should be similar to that of the archival speakers. The analyses were based on a total of 2.72 hours of archival and 2.68 hours of present-day spontaneous speech material. The disfluency phenomena (of the “speaker’s uncertainty” type) studied were as follows: filled pause, repetition, filler word, prolongation, pause in the word. The error types were: wrong word, contamination, restart, anticipation, perseveration, false start, grammatical error, tip of the tongue phenomenon, slip of the tongue.

The aim of the comparative study was to see if there were any differences between speakers of sixty-odd years ago and today’s speakers with respect to the planning and execution processes of spontaneous speech. It was hypothesized that there would be dissimilarities in the frequency of occurrence of the individual disfluency phenomena, witnessing differences in speech planning processes, hence interpretable as signs of linguistic change. The archival material contained a total of 446 uncertainty-type disfluency phenomena, whereas that of the present-day speakers contained 1346 comparable instances. The figures include the number of silent pauses, too. The number of errors also increased in today’s speakers as compared to those of the archival material: the latter exhibited 122 whereas the former exhibited a total of 408 instances. The most spectacular change occurred in the occurrence of filled pauses. They hardly ever occurred in the speech of the archival speakers, a mere 18 filled pauses were found in the whole material studied. Present-day speakers produced 636 filled pauses in their spontaneous speech of the same length. One possible reason might be that spontaneous speech
is characterized by longer and more complex utterances today than it was in the fifties, as the recordings suggest. More complex constructions put a heavier load on the planning processes, and this may underlie the increase of the number of filled pauses. Another possible explanation is that the archival speakers produced far more filler words than today’s speakers did, especially favoring the phrase *azt mondja* ‘it says’, rather than uttering hesitation noises (filled pauses). Wherever the subjects were uncertain as to what to say next, or there occurred a temporary disharmony in their speech planning processes, they most often uttered some filler words. The proportions of filler words in today’s speakers are similar to those in the archival speakers (273 and 296 occurrences, respectively). A smaller but still mathematically significant difference was found in the case of repetitions (42 in the archival material, and 207 in the present-day material). Similarly, a significant difference was found in the occurrence of prolongations (84 and 140, respectively).

Archival speakers committed a total of 130 errors in their 2.7 hours of spontaneous speech, whereas present-day speakers produced 390 disfluency phenomena of the error type in the same amount of time. The number of errors attested in the present-day corpus is exactly three times as many as found with speakers of the forties or fifties. Statistical analyses confirmed that there were significant differences between the two corpora in four error types: restarts, anticipations, grammatical errors, and slips of the tongue. The discrepancy is conspicuously large in grammatical errors: 12 were attested in the archival material and 111 in the new one. The number of anticipations and slips of the tongue are seven times as large as they used to be, modified restarts and replacements were multiplied by three, and wrong words and simple restarts increased to twice the number they had in the old material.

In the archival corpus, the speakers produced 0.74 errors and 2.72 uncertainties per minute, a total of 3.46 attested cases of disfluency in a minute. Present-day subjects produced 2.53 errors and 8.17 uncertainties, a total of 8.17 attested instances of disfluency each minute. In view of the fact that the speech rates of the two populations were nearly identical, such differences in speech planning and execution give food for thought. In spite of the fact that significant differences were not found for each and every disfluency phenomenon, it is still noteworthy that the number of occurrences was larger in present-day speech in all categories, even if to different extents. The spontaneous speech of present-day speakers, then, is definitely far more spasmodic than it used to be.

Analyzing the proportions of disfluency phenomena corpus-internally, it was found that while earlier on the strategies of resolving disharmony primarily involved filler words and prolongations (in 80% of the cases), today the main exponents of such strategies are filled pauses and repetitions (in 84.6% of the cases). It appears that all this shows a change of habits in language use. With respect to errors, archival speakers exhibited more wrong words, false starts and instances of the tip of the tongue phenomenon, while present-day speakers exhibited grammatical errors, slips of the tongue, and anticipations more often than any other error types.

In present-day speakers, the relatively large number of grammatical errors, the frequent occurrence of anticipations, and a strong increase in the number of filled pauses show that they have difficulties in finding an adequate linguistic form for the expression of their thoughts. On the other hand, the speech planning processes of the archival speakers proved to be more hampered by difficulties in activating their mental lexicon than in creating the appropriate grammatical form for what they wanted to say.

Potential changes in articulation rate and speech rate were explored on the basis of the Hegedűs Archives, the BEA database [10] containing contemporary standard speech samples, and recordings of subjects from a village in Nógrád County [15]. The results confirmed that there are significant differences between the archival speakers and both present-day
populations (village-dwellers and inhabitants of the capital) in articulation rate (mean values: 9.9 sounds/s, 11.1 sounds/s, and 12.6 sounds/s, respectively). In speech rate, there was no difference between archival speakers and today’s village people, but both groups differed significantly from the Budapest speakers (7.9 sounds/s, 8 sounds/s, and 9.7 sounds/s). The comparison showed that there were differences in pause durations across all three groups. Silent pauses were the longest in recordings of the Hegedűs Archives (mean: 653 ms), shorter than that in the speech of villagers (mean: 596 ms), and the shortest in that of present-day Budapest speakers (mean: 518 ms). Archival speakers paused primarily at phrase boundaries (in 65% of the cases), as opposed to today’s speakers who only paused at those boundaries in 52% of all cases. The study appears to confirm a relative acceleration of speech rate over the decades, at least with respect to those groups.

Temporal properties of the speech of fourteen young subjects of the Archives aged 10 to 12 and 14 to 16, respectively, in spontaneous speech versus storytelling [16]. The results show that storytelling is slower than spontaneous narratives are, and that the speech rate of older children is slightly higher than that of younger children. Pauses occur more often in narratives than in storytelling. Both articulation rate and speech rate are somewhat faster in these children than in adult speakers of the Archives.

Recordings of the Hegedűs Archives also give us an excellent chance for studying the various genres of speech (tales of various sorts, spells, superstitions, descriptions of folk customs, spontaneous narratives, etc.). Menyhárt [17]’s investigations showed that the speakers’ speech rate was the fastest in relating fairy tales, and the slowest in spontaneous narratives. Pitch range, too, was the widest in fairy tales. The lowest fundamental frequency was found in spontaneous speech, and the highest in tales about animals. Archival speakers produced the briefest pauses in ritual(ized) texts; the average durations of silent pauses were nearly equal in tales and in spontaneous narratives.

4. Outlook

In the second half of the twentieth century, large speech databases were created in a number of languages and for a variety of purposes. However, there are few if any corpora that are instrumentally analysable but contain material coming from the mid-twentieth century and that have been created by linguists primarily for the purposes of phonetic studies. The main significance of the Hegedűs Archives lies in the fact that it makes it possible for us to study the properties of spontaneous and semi-spontaneous speech of over 70 years ago, to compare them with present-day speech, and that the recordings can also be studied in sociolinguistic, dialectological, and ethnographic terms. The contents of the speech materials are part of our cultural heritage. The Hegedűs Archives can be taken to be an up-to-date database as it consists of a searchable written database and an organised speech material, making all sorts of utilisations easy to perform. Full annotation of the whole material of the Archives, however, remains a task for the future.

References


