The LDS General Conference Recordings Corpus

David Bowie[1]* & Kjerste Christensen[2]

[1]University of Alaska Anchorage, United States
[2]Brigham Young University, United States

Abstract. The LDS General Conference Recordings Corpus is an archive containing semi-extemporaneous English-language sermons delivered by leaders of the Church of Jesus Christ of Latter-day Saints, with several individuals appearing repeatedly across spans of decades within the corpus (which covers 1936 to the present). While not originally designed for linguistic research, the corpus offers an opportunity to observe linguistic variation and change across the adult lifespan. Its value for language and aging research lies in its temporal depth, consistency of genre, and public prominence of its speakers, which allows for reliable biographical contextualization. Despite limitations such as lack of linguistic annotation, limited demographic diversity, and access constraints, the corpus has already supported studies of age-related phonological change and holds further potential for exploring lifespan trajectories. It serves as a compelling example of how found corpora can advance language and aging research.

Keywords. archives, corpora, language and aging research, recordings, sermons

© David Bowie & Kjerste Christensen

1 Introduction

The LDS[2] General Conference Recordings Corpus[3] is a partially publicly available set of recordings of semi-extemporaneous sermons delivered by leaders of the LDS church

---

*. Corresponding author, david.bowie@alaska.edu

OPEN ACCESS

2. The *LDS* in the name of the corpus (and as used elsewhere in this paper) refers to the Church of Jesus Christ of Latter-day Saints, one of the "new religious movement" churches that emerged in the United States during the Second Great Awakening period. This church is perhaps most widely known by the label *Mormon*, but *LDS* is the preferred term among group members (Bowie and Petersen 2018).

3. The corpus is accessed through the LDS Church History Catalog (Church History Library 2025), following the steps outlined in Section 2.5.

from the mid-1930s to the present. It was not originally developed as a linguistics corpus, but rather primarily as a religious corpus and secondarily as a historical one that was digitized and placed online by the LDS church's Church History Library. This places it in the category of what are sometimes called "unprincipled corpora," a potentially misleading term, since the term *unprincipled* refers purely to the way that the texts are collected with regard to intentionality for specifically *linguistic* research. There is a general recognition that these types of corpora can be valuable resources for linguistic research (see, for example, Curzan and Palmer 2006). The "unprincipled" nature of this corpus does mean that there is no linguistic annotation, limiting the sorts of analyses that can be conducted using it. It is, however, particularly useful for language and aging research simply because it consists of recordings made over the course of decades, often featuring the same speakers repeatedly across long periods of time, which allows for (retrospective) tracking of individuals' linguistic production with regard to age.

We present this corpus here to provide the wider research community information about it, of course, but also in part to highlight a simple fact: That corpora exist "in the wild," so to speak, and these can be successfully used for linguistic analysis. Such corpora will not always be well suited to language and aging research in the way that the LDS General Conference Recordings Corpus is, of course, but they can be creatively repurposed to provide linguistic insights, including some that corpora developed specifically for linguistic analysis cannot (or at least cannot yet) give us.[4]

2 Corpus description

2.1 History and content of the corpus

The LDS church's first general churchwide conference occurred in 1830, the year the church was incorporated, and they have been held nearly every April and October since 1838. These conferences have been broadcast over radio since 1924 and television since 1949. Recordings survive—and, thus, the content of the LDS General Conference Recordings Corpus dates—from 1936 onward, although the recordings we have from the first few years of the corpus are fragmentary.

The corpus consists of two collections (one audio-only, one audiovisual) of recordings of sermons, prayers, items of church business, and songs from each conference. The focus for this description is the sermons, however, because they are the content that makes the corpus most applicable to language and aging research. The sermons in the corpus are all in English,[5] and were all delivered simultaneously to a live and a broadcast audience from one of three locations, all within a city block of each other in Salt Lake City, Utah, USA: Conferences through 1999 were all delivered in the Tabernacle except for a few during World War II that were held in the smaller adjacent Assembly Hall due to wartime restrictions on meeting sizes, and all conferences held since 2000 have been held across the street in the Conference Center.

---

4. This is, of course, not the only "unprincipled" corpus that has been used to gain insights into language and aging—see, for just a few earlier examples, Trudgill's (1983) and Yaeger-Dror's (1993) use of recorded music, and van de Velde, Gerritsen & van Hout's (1996) and Harrington, Palethorpe & Watson's (2000) use of radio broadcasts.

5. This is not absolutely true, since a handful of these sermons were delivered in other languages in the mid-2010s. This makes up such a small portion of the corpus, however, that it is left to the side in the description here.

The precise format of the conferences has varied somewhat over the years, but a typical conference might last 10 to 14 hours over two or three days, featuring 25 to 35 speakers each delivering semi-extemporaneous sermons of seven to 20 minutes in length, and a few occasionally going longer than that. These sermons are given by individuals selected from among leaders of the LDS church, particularly those who are "general authorities and officers"—that is, holding positions at a general churchwide level as opposed to leading a particular congregation or set of congregations.

Because there is a limited pool of speakers to choose from, then, the same individuals can appear in the corpus multiple times over the course of decades. This leads to some individuals appearing frequently over time, because every general conference includes at least one sermon—health permitting—from each of the individuals holding one of the fifteen highest leadership positions in the LDS church, which are lifetime appointments.

As a result, many of these individuals appear in the corpus annually (or nearly so) over long periods of time—for example, LeGrand Richards appears in the corpus from 1938 to 1983, and Gordon B. Hinckley appears from 1958 to 2008. The corpus's application to language and aging research, therefore, is clear, because it provides researchers the ability to track these individuals' linguistic production over the course of decades of their lives[6] as they navigate the challenges of aging—which is occasionally made explicit by speakers in these sermons, as, for example, David B. Haight did when opening his address in October 1995 (when he was 89 years old) with the line "As some of us get older, we slow down, so you'll have to tolerate us a little."

2.2 Style and audience

Before going further, it is worth mentioning that the LDS church does not have any sort of seminarial training for its leaders—LDS church leadership, at both the local and general levels, is chosen from the lay membership. A side effect of this is that those who speak in LDS general conferences have received no special training in the creation or delivery of sermons. There are, of course, traditions and patterns for the delivery of sermons in the LDS context, but these are acquired through exposure rather than direct instruction (Knowlton 1991; Gilkey 1994; Betts 2019; Hale 2024).

LDS sermon style historically comes from a "low church" tradition, which is to say that ritual and liturgy are deemphasized, but references to scripture[7] and personal experience are emphasized. Over time it has further developed into a form in which occasional displays of emotion—including audible cues in a speaker's voice—are appreciated, but in general the speech style is relatively formal (Gilkey 1994), with a cadence that Knowlton (1991, 26–27) describes as "slow and evenly pulsed" with "tonal contours [that] become leveled . . . , dropping down at the end."

The audience for these sermons plays a role in their delivery, as well. Though they are broadcast using various forms of mass media (radio, television, internet) that are accessible by anyone with an interest, there is a live audience, and they are primarily

---

6. We would like to acknowledge Bill Eggington as the one who pointed out this facet of the corpus to one of the authors in a conversation in 2002, in the context of a discussion of doubts about the apparent-time hypothesis, and the lack of ways to adequately test its reliability. This is why some early papers using this corpus focused on real- versus apparent-time analyses (e. g., Bowie 2005) rather than focusing on aging directly.

7. LDS church scripture includes the Bible, along with a set of other books that are considered canonical by the church.

directed toward fellow members of the LDS church. Thus, there are norms that are adhered to by the speakers that serve as models of LDS sermon style for the general membership of the LDS church (Knowlton 1991), while the general membership of the church has expectations about what is acceptable for LDS sermon style (Gilkey 1994). In this way the speaker and audience together form a community of practice (Wenger 1998) that mutually reinforces the existence of LDS sermon style and norms surrounding the way it is performed (Hale 2024).

The tone of the addresses tends to be relatively formal, even staid, though there are some points in many of the sermons that provide a break from this (e. g., the occasional use of humor; see Hale 2024). This is reflected in things such as sentence and discourse structure in most of the sermons, and the brief examples below are given to provide an idea of the sort of tone one finds in them, as well as, to a very limited extent, the use of in-group language.

> My brothers and sisters, it is a wonderful privilege to be able to attend this conference with you today. I thank the Lord for membership in this great church. When Jesus was among men, he said that the gospel was the pearl of great price, and one seeking costly pearls would sell all that he had in order that he might acquire the pearl of great price. I thank God that — that I possess this pearl of great price through having membership with you in this great church. And then Jesus indicated that if we would seek first the kingdom of heaven and his righteousness that all other things would be added thereunto. Not that we seek last, that we give him but the crumbs that fall, as it were, from the master's table, but that we seek first the kingdom of God and his righteousness. And I thank the Lord that there are so many faithful Latter-day Saints who are seeking first the kingdom and they are enjoying the gifts and the blessings of the Lord which I think are above all other riches in this world. (LeGrand Richards, October 1945, 0:00–1:33)

> Man has an innate urge to worship. In early times God spoke to Israel: "I am the Lord thy God, which have brought thee out of the land of Egypt, out of the house of bondage. Thou shalt have no other gods before me." There is a profound truth in the doctrine that runs all the way through sacred history that no man can worship more than one God. To worship one God is to have one supreme loyalty in one's life. If we have a consciousness of one God, the eternal father, we would have a consciousness of one world and one mankind under God, all being brothers. What causes people to have the urge to worship? There seems to be something inborn into the soul of man which causes him to seek communion with God. In the Book of Job it is put this way: "But there is a spirit in man: and the inspiration of the Almighty giveth them understanding." This statement appears to be an allusion to man's creation. By this spirit he becomes capable of understanding and reason, and consequently of discerning divine truth. By this spirit he comes to know God. (Howard W. Hunter, April 1970, 11:17–12:55)

> We choose to be steadfast and immovable in our faith because of the promises of eternal glory, eternal increase, and continued family relationships

in the celestial kingdom. We love our families and know that our greatest joy and peace come to us as we watch each family member face the tests of life and make righteous choices to overcome the world. Occasionally I place my hands on both sides of the face of one of my children or grandchildren when they are doing something that will bring immediate or long-term harm to themselves in the process. I look deeply into their eyes and carefully explain to them how much they are loved and cherished. Then I describe the harm that could result from the actions they have chosen. I can envision the Savior holding our faces between his hands and pleading with each of us individually to remain steadfast and immovable and faithful to the God who made us. (Mary Ellen W. Smoot, October 2001, 1:02–2:24)

2.3 Limitations (and advantages) of the corpus

Because of the nature of the corpus, it presents a researcher with a few limitations. First and most obviously, the corpus does not lend itself well to analyses of linguistic change and variation with regard to social factors other than age. So, for example, the corpus is by its nature limited to recordings of practicing members of the LDS church, which is important given that a majority of the recordings in the corpus are of individuals who use Wasatch Front English (often referred to in the literature less precisely as "Utah English"), and we know that Wasatch Front English speakers vary linguistically by both religious affiliation and religiosity (Di Paolo 1993; Baker-Smemoe and Bowie 2015). Further, given the leadership roles filled by the speakers in the corpus, all of them hold a similarly high social status in the community of practice that they and the vast majority of their audience participate in. In addition, there are no female speakers in the LDS General Conference Recordings Corpus before 1984, and because the leadership of the LDS church skews very heavily male, even from that point there have been relatively few female speakers, and even fewer who appear multiple times. There is also not a wide range of stylistic variation in the corpus—the vast majority of the corpus is made up of sermons, and so the style is consistently rather formal, given the in-group norms for preaching described above.

In addition, this corpus has not yet been fully transcribed in a way that easily supports analyses of linguistic production across the lifespan, both in terms of a lack of linguistic annotation as well as the nature of the (unannotated) transcriptions that do exist. There is a separate corpus of transcriptions of these recordings, the LDS General Conference Corpus (Davies 2011–). However, that corpus uses the published transcripts of the sermons given in the LDS church's general conferences, which do not always follow the sermons as they were actually delivered. This is not only because those transcripts are edited for fluency and style, but also because the speakers in the conferences are given the chance to edit what they said, because the published written form is considered the definitive record by the church—and so the authors sometimes take the opportunity to make major changes.

However, these limitations also present users of this corpus with an obvious opportunity: With so much social uniformity among speakers aside from year of birth, it is possible to focus on issues of language and aging without having to do as much work to filter out possible sociolinguistic confounds as might be needed with a corpus that contains a wider range of social factors represented in its speakers. it is, after all, not necessary—perhaps better, not possible—for a single corpus to be all

things to all researchers, and the result is that this corpus is particularly well set up to investigate interindividual and intraindividual linguistic differences with respect to age.

2.4 Metadata, documentation, and preservation

Since this corpus was developed primarily for historical and religious research, the metadata connected with the corpus reflects that. This means that it does not contain much of the structured metadata that one expects in a linguistics corpus — for example, there isn't even a reliable count of the number of words contained in it.[8] However, the metadata that is available still does provide useful information.

All of the recordings in the corpus are provided with the date the recording was made, and then the speaker(s) included in the recording; most, though not all, of the audio recordings also include timestamps to ease finding a particular speaker. In most cases, notes are given to indicate whether a speaker was recorded conducting items of church business or offering a prayer; the sermons, as the default, are left unmarked. There are also occasional notes about the recordings, generally indicating problems; for example, in the audiovisual collection, the 6 April 1956 recording of David O. McKay indicates that only two-thirds of that sermon was recorded on video. (Though it is not mentioned in that entry, the audio collection contains the entirety of the speech.)

This is, obviously, a very small amount of metadata, and so it might seem that this does not lend itself well to linguistic research. However, it turns out that a good deal of what might be thought of as *indirect* metadata is available, and this information makes the available metadata, and thus the corpus as a whole, quite useful for linguistic research, and particularly for language and aging research. Since the speakers are religious leaders, once one has a speaker's name it is trivial to find sociolinguistically useful information about them ranging from date of birth to their educational and employment background to at least the general outlines of a residence history.[9] With this additional information that is not provided directly by the corpus's metadata but which that metadata can be connected to, it is straightforward to connect social variables such as age and region of speaker to linguistic tokens collected from individual sermons.

Finally, as already mentioned, this corpus was not created by linguists to be a corpus in the way the field usually defines a corpus. Consequently, linguists are not involved in its ongoing accessibility and preservation. However, the LDS church considers them to be important religious texts, and thus preserving and making them available is a high priority, which means that there is no reason to expect that they will become unavailable in the foreseeable future. Further, while the only available digitization of first-generation recordings of these conferences is this corpus, there are other depositories (e. g., Brigham Young University's Harold B. Lee Library) housing second-generation copies of large portions of the corpus, which function as effective backups for it, albeit not as easily accessible.

---

8. The closely related text-based LDS General Conference Corpus (Davies 2011–) says that that collection — which covers a longer timespan, from 1851 to the present, than the LDS General Conference Recordings Corpus — contains about 25 million words, but those are edited (and in some cases heavily edited) transcriptions.

9. As of this moment, Wikipedia functions as a reliable resource for this sort of information, though there are a number of other sources available, both online and in print.

2.5 Availability and permissions

The corpus of recordings has, as mentioned above, been digitized and placed online by the LDS church's Church History Library, and is available in two different collections covering different but overlapping timespans: an archive of audio recordings (titled "General conference, digital audio, 1936–2012") and an archive of video recordings that include audio (titled "General conference, digital video, 1949–2024", though since collection appears to be ongoing for this collection the end year in the title is likely to change). Both collections can be accessed through the library's catalog (Church History Library 2025) and entering in the collection number (AV 3501 for the audio-only collection, or AV 3500 for the audiovisual collection).

The content of these two collections differ in a number of important ways. First, for many of the years that they overlap, the audio-only collection is more complete than the audiovisual collection. The latter, especially in its earliest years, sometimes contains only selected sermons from a given conference, while the audio-only collection nearly consistently contains complete or nearly-complete recordings for those years. (Some of the earliest years of the audio-only collection are also fragmentary, but in all cases the audio collection contains every recording that is known to have survived.)

However, more of the audiovisual collection is freely accessible without authentication than the audio-only collection. This is due to an interaction between the way the collections are divided up and copyright restrictions. The corpus contains musical performances alongside the sermons, and not all of the songs that have been performed over the decades have been licensed by the LDS church for online streaming (which is, effectively, the way the recordings are distributed). In many years the audiovisual collection has individual sermons segmented out, while the audio-only collection is more likely to be presented as an entire session, and thus if, say, a particular sermon was given adjacent to a musical performance that isn't licensed for streaming, both of them are subject to authentication for access because they are in the same recording.

The entirety of both collections of the corpus, though can be accessed via an authentication system. Authentication is currently done through a login using a "church account," which requires registration with the LDS church to use certain features of their websites. It is primarily intended for members of that church, but is also available to those unaffiliated with the LDS church. Account registration is free, but does still present some level of a barrier to full access to the corpus.

Before continuing on, it should be noted that these are recordings of human beings, and therefore their use is human subjects research. However, most of the individuals contained in the corpus are deceased, and all of the recordings were broadcast publicly with the knowledge of the speakers that they would be recorded and archived. Because these recordings were initially made with no expectation of anonymity or privacy on the part of the speakers, researchers (whether in linguistics or other fields) who have used this corpus in the past have not needed to receive clearance from human subjects research review boards. Before undertaking a project using this corpus (or a similar one), however, one should always consult with one's own review board to see if local regulations require oversight.
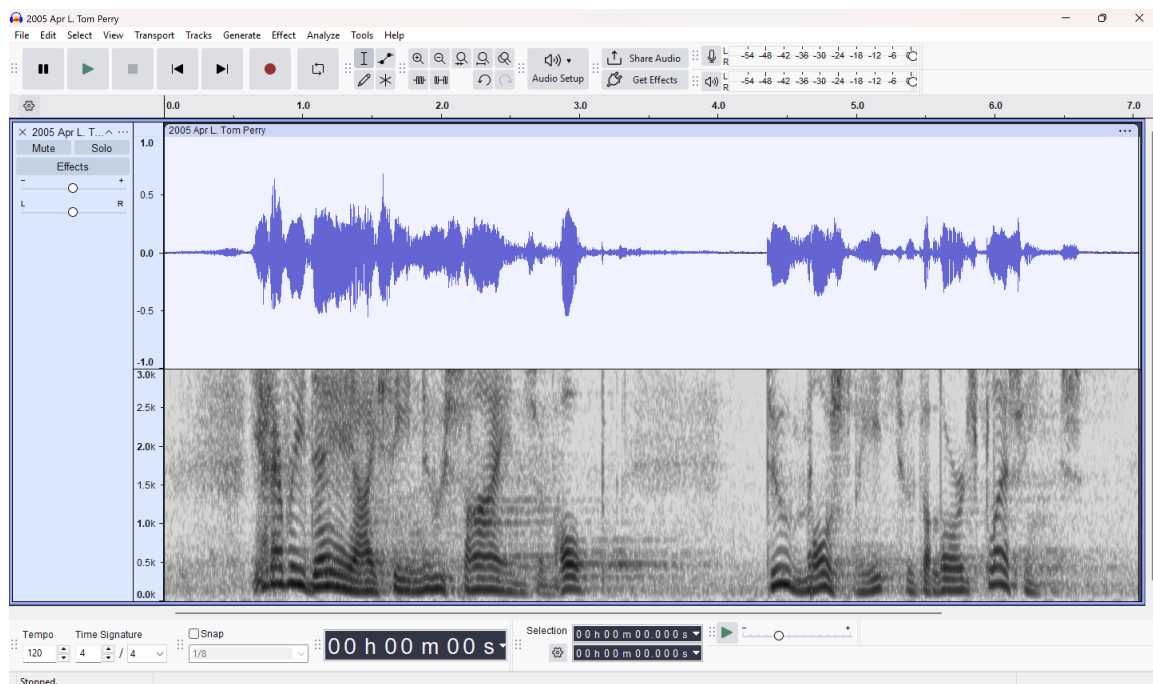
Figure 1: An oscillogram and spectrogram of audio extracted from the audio collection of the LDS General Conference Recordings Corpus (L. Tom Perry, April 2005)
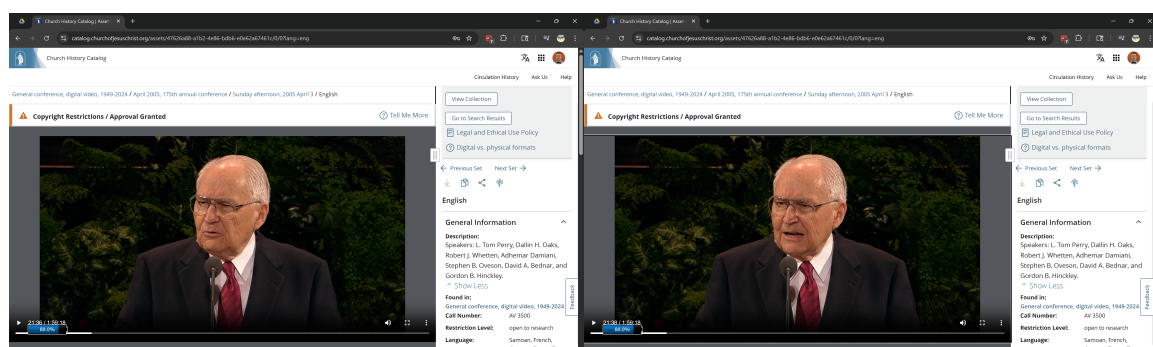

Figure 2: Two video stills captured from the video collection of the LDS General Conference Recordings Corpus (L. Tom Perry, April 2005)

3   Examples of audio and video in the corpus

To provide a more concrete view of the content of the corpus and what can be done with them, Figures 1 and 2 show examples of the content. Both of them are from a specific address delivered in April 2005 by an individual who appears multiple times in the corpus, L. Tom Perry.[10]

---

10. Perry's first appearance in the corpus occurred in 1972 (when he was 50 years old), and he spoke in nearly every conference contained in the corpus from then until his death in 2015 (at age 92).

Figure 1 is a few seconds of Perry speaking (saying "Yet every day I see new signs of hope, due mostly to the Lord's blessings"), extracted from the audio collection.[11] Figure 2 shows two stills from the video collection, taken during the same utterance, using the built-in video player on the Church History Library website; the one on the left shows Perry producing the vowel in the word *due*, and the second shows him producing the first vowel in the word *blessings*.

The video recordings are filmed from a number of different angles, but primarily from the front or a three-quarters profile of the speaker, with an occasional full profile. This generally provides a good view of such things as lip shape and head movement. Hand gestures, however, are only visible if they are around chest height or higher.

4   Using the corpus

4.1   Possible applications

The types of linguistic analysis that this corpus is well suited to are limited by its nature. Since, as already mentioned, the sermons recorded in it are *semi*-extemporaneous rather than fully extemporaneous, this corpus is likely not very useful for analyses of syntactic or morphological change and variation in individuals across the adult lifespan, since even partially pre-prepared or pre-written sermons would be expected to be heavily leveled toward standardized forms. It could, however, presumably be used to investigate phenomena ranging from discourse features to prosody, though we are unaware of any such studies using this corpus to those ends. It also lends itself well to work on lexical and semantic change, though the studies that have used the corpus for that have not focused on age-related effects (e. g., Madsen and Corey 2023; Teerlink 2024; Williams 2024).

On the other hand, the corpus is very well suited to the study of phonetic and phonological variation and change, and it has in fact been used to do precisely that, both without (e. g., Bowie 2012, 2017) and with (e. g., Bowie 2019, 2021) a focus on aging. There are, however, some necessary cautions when conducting such analyses. First, the recordings in the corpus are in MP3 and MP4 formats, which are lossy, and so cannot be relied upon for things like formant extraction. Further, some of the original sources of the digitized recordings have flaws that could skew measures of the sound waves themselves (e. g., a warped or otherwise damaged master resulting in unavoidable auditory issues on the resulting shellac and vinyl pressings), which may or may not be easily detectable simply by listening to them. Also, since recording technologies and methods have changed over time, instrumental analyses are subject to a possibly insurmountable number of confounding issues even for those who have access to lossless versions of the recordings (Bowie 2024). And finally, these are, essentially, field recordings of speeches delivered in very large rooms (e. g., the Tabernacle, where most of these recordings were made, seated about 7,000 people) with a noticeable amount of reverb and echo, plus occasional audience noise mixing with the audio of the speaker[12] as well as—especially in the earlier years of the

---

11. Though a spectrogram is given in Figure 1, the formants shown in it should are not fully reliable, for reasons described in Section 4.1.

12. Audience noise is infrequent, however, in large part because there is no call and response tradition in LDS worship.

corpus — some clipping. Therefore, it is safest to conduct analyses of phonetic and phonological variables that do not require instrumental analysis.

Whatever variables are chosen, however, the selection of individuals to sample from the corpus is relatively straightforward, because all of them are identified by their actual name, and a large amount of biographical information is easily available for most of them due to their status as public religious figures. Therefore, it becomes a simple matter of identifying which speakers appear repeatedly over time in the corpus and also fit any desirable demographic characteristics (e.g., limiting the sample to speakers who all come from the same dialect region), and using their addresses as a data source.

4.2 Existing work using the corpus

As already mentioned, this corpus has already been used for a number of studies in linguistics and other fields. Within linguistics, several of them have focused on language and aging research, providing information on linguistic change and variation across the adult lifespan (Bowie 2005, 2010, 2011, 2015, 2019, 2021). These studies have given us some insight into the range of phonological variation that is possible for individuals to produce throughout the decades of adulthood from middle to old age, and confirmation that there is no single linguistic trajectory that individuals, even similarly situated individuals, follow as they age through adulthood.

To provide an example of the work that can be produced with this corpus, consider Bowie's (2015, 2021) findings regarding syllable-initial (wh). This linguistic variable, also known by labels such as the *wail-whale* or *wen-when* merger, is the result of the in-progress (though nearing completion) historical change in English where the Modern English reflex of the Old English [hw] (as in the Old English word *hwenne*, Modern English *when*) is losing its initial [h] and thus merging with the Modern English reflex of the Old English [w] (as in Old English *wenn*, Modern English *wen*).

The merger of [hw] into [w] has progressed to completion throughout much of the English-speaking world, but among the places where it is still in progress is the Wasatch Front region of the United States, which has Salt Lake City as the major metropolitan center. A large portion of the individuals in the LDS General Conference Recordings Corpus grew up in the Wasatch Front region, and so are surrounded by (wh) as a change in progress that, at least at the time the speakers discussed here were being recorded, was not socially salient.[13]

Table 1 shows several different individuals in the corpus (all from the Wasatch Front of Utah, and so speakers of that variety of English), arranged by year of birth, with the percent of the time each of them produced the conservative form — that is, the [hw] — of the (wh) variable in the years divisible by five that they appear in the corpus. This is a somewhat complicated table, but two things are clear. First, looking at the table with an eye toward apparent time provides some evidence that (wh) is, in fact, changing over time in this region, since speakers born more recently are more likely to produce the [hw] variant at lower rates.

---

13. The variable had a burst of pop-culture salience when it was used as a running gag in a 2006 episode of the animated series *Family Guy*, to the point that that episode was referenced shortly thereafter in an advertisement for the snack cracker Wheat Thins (which can be pronounced with an initial [hw] in places where the merger with [w] is not yet complete). As far as we can tell, though, the salience of the feature has not continued, and it does not come up in Utahns' linguistic metacommentary about Utah English accents.

Table 1: Rates of retention of [h] in syllable-initial (wh) in percent (0 to 100, with darker cell shading for higher values), by speaker and year of recording (adapted from Bowie 2015 Table 3 and Bowie 2021 Table 4.4)

Speaker	Birth year	19												20		
		40	45	50	55	60	65	70	75	80	85	90	95	00	05	10
JRC	1871	37	40	53	40	18										
DOM	1873	52	48	74	49	55	57									
JFS	1876	33	50	43	17	44	67	47								
HBB	1883		8		23	13	17	13								
LGR	1886	72	64	94	47	73	52	50	43	38						
JLW	1893	15	28	58	14	28										
SDY	1897		37	53	30	36	71	43	44							
MEP	1900		52	36	51	39	55	52	65	30						
MRH	1902		32	37	20	70		47								
DBH	1906							13	42	12	48	12	35	15		
RLE	1906	31	30	50	33	20	18	30								
GBH	1910					56	41	61	69	48	65	47	30	51	41	
BKP	1915						46	29	62	29	13	21	27	13	13	22
BRM	1915			48	67	66	69	57	36	51	33					
MJA	1915							8	6	34	12	14				
JBW	1917								4			7	0	7	0	
MDH	1921					43	37	13	37	10						
TSM	1927				23		9	16	10	14	5	5	7	7	3	13
MRB	1928						13			17	12	27	47	7	13	13

Second, and more intriguing, is that change across adulthood—that is, across years of recording for any given individual—does not monotonically change in one direction or the other, and there is often significant variation from year to year.[14] This is particularly interesting given previous research that has found that speakers, if they change their phonetic production during middle to old age, may change in any direction but are likely to simply change in the same direction as the community as a whole (Sankoff, Blondeau, and Charity 2001; Sankoff 2018, 2019). Most studies looking at change across the lifespan have looked at only two or three points in time, however, and the data here shows that a more fine-grained look at linguistic behavior across the lifespan provides evidence of a more complex—and really, more interesting—pattern, where there is movement up and down in terms of frequency of one variant or the other. Thus, while some of the speakers who show a clear trend in this variable across middle and old age do move in the direction of the change in progress (e. g., LGR, BKP) and others show the change toward a more conservative form that Sankoff (2018, 2019) called "retrograde change" (e. g., MRH, MJA), most speakers' behaviors across age do not fall into an easily classifiable shape. Thus, the fact that this corpus allows us to look at linguistic behavior across age with more granularity than most studies have been able to achieve allows us to conclude that speaker behavior in middle and old age is more dynamic than often reported.

5   Conclusion: The utility of so-called unprincipled corpora

There is, however, much more that could be done using this "unprincipled" corpus, as well as other corpora made up of what might be called "naturally"—that is, natural in the sense of not having been collected and classified by *linguists* specifically—collected language (see Curzan and Palmer 2006).

For example, corpora with individuals reappearing across time like this one allow us to know the aging experience of groups of people whose divergent experiences of aging do or do not include experiencing age-associated cognitive and physical disorders, allowing us to approach our study of lifespan linguistic change and variation in a way that fills a gap identified by Pichler, Wagner, and Hesson (2018). In addition, similarly built corpora with a different focus (say, to give one possibility, a corpus of recorded legislative debates and speeches) could allow a better understanding of the ways that issues like genre and register are negotiated across the lifespan.

These are just two possibilities, but there are, of course, others—the important thing is that for those of us interested in language and aging research, there are corpora all around waiting to be used to advance our understanding, and it is just a matter of identifying them and figuring out how they can best be used.

6   Conflict of interest

The authors have no conflicts of interest to declare.

---

14. This brief discussion ignores complications such as differing numbers of tokens from year to year, nor are statistical measures described here. Readers interested in that information are referred to the articles by Bowie (2015, 2021) in which these findings originally appear.

Funding

Ethics statement

The description of the corpus did not involve human subjects research. The previously published research using the corpus that was summarized in this paper did not require human subjects review, per the University of Alaska Anchorage's Institutional Review Board, because the recordings that were analyzed had been publicly broadcast with the consent of the speakers.

References

Baker-Smemoe, Wendy, and David Bowie. 2015. "Linguistic behavior and religious activity." *Language & Communication* 42:116–124. https://doi.org/10.1016/j.langcom.2014.12.004.

Betts, Stephen Thomas. 2019. "'General conference talk': Style variation and the styling of identity in Latter-day Saint general conference oratory" [in English]. Master's thesis, Brigham Young University.

Bowie, David. 2005. "Language change over the lifespan: A test of the apparent time construct." In *Selected papers from NWAV 33,* edited by Suzanne Evans Wagner, 45–58. Philadelphia, Pennsylvania: University of Pennsylvania Working Papers in Linguistics.

———. 2010. "The ageing voice: Changing identity over time" [in English]. In *Language and identities,* edited by Carmen Llamas and Dominic Watt, 55–66. Edinburgh, Scotland: Edinburgh University Press.

———. 2011. "Aging and sociolinguistic variation." In *Language, culture, and the dynamics of age,* edited by Anna Duszak, Urszula Okulska, and Richard J. Watts, 29–51. Berlin, Germany: Mouton de Gruyter. https://doi.org/10.1515/9783110238112.2.29.

———. 2012. "Early trends in a newly developing variety of English." *Dialectologia* 8:27–47.

———. 2015. "Phonological variation in real time: Patterns of adult linguistic stability and change." In *Language development: The lifespan perspective,* edited by Annette Gerstenberg and Anja Voeste, 39–58. Amsterdam, The Netherlands: John Benjamins.

———. 2017. "Early development of the Western Vowel System in Utah." In *Speech in the western states: Volume 2: The Mountain West,* edited by Valerie Fridland, Alicia Beckford Wassink, Tyler Kendall, and Betsy E. Evans, 83–105. Durham, North Carolina: Duke University Press. https://doi.org/10.1215/00031283-4295211.

Bowie, David. 2019. "Individual variation in the development of the Western Vowel System of Utah." Edited by Annette Gerstenberg and Torun Lindholm. *Linguistics Vanguard* 5:Article 20180020. https://doi.org/doi.org/10.1515/lingvan-2018-0020.

———. 2021. "Individual and group trajectories across adulthood in a sample of Utah English speakers." In *Language variation and language change across the lifespan: Theoretical and empirical perspectives from panel studies,* edited by Karen V. Beaman, Isabelle Buchstaller, Isabelle Buchstaller, and Suzanne Evans Wagner, 101–118. New York, New York: Routledge.

———. 2024. *Building and working with a panel corpus across decades of speakers' lives (and decades of recording technologies).* Poster presented at CLARe6, Tübingen, Germany, April 11. Tübingen, Germany.

Bowie, David, and Boyd J. Petersen. 2018. *Group naming practices and sociocultural borders in the Mormon Culture Region.* Paper presented at Language and Borders: Rethinking Mobility, Migration, and Space, Bristol, England, March 26.

Church History Library. 2025. *Church History Catalog (accessed May 9, 2025).* Salt Lake City, Utah: The Church of Jesus Christ of Letter-Day Saints. https://catalog.churchofjesuschrist.org/.

Curzan, Anne, and Chris C. Palmer. 2006. "The importance of historical corpora, reliability, and reading" [in English]. In *Corpus-based studies of diachronic English,* edited by Roberta Facchinetti and Matti Rissanen, 17–34. Bern, Switzerland: Peter Lang. https://doi.org/10.3726/978-3-0351-0269-7/3.

Davies, Mark. 2011–. *LDS general conference corpus.* Text corpus. https://www.lds-general-conference.org/.

Di Paolo, Marianna. 1993. "Propredicate *do* in the English of the Intermountain West." *American Speech* 68:339–356.

Gilkey, Carolyn Flatley. 1994. "Verbal performance in Mormon worship services." PhD dissertation, University of Pennsylvania.

Hale, Adrian. 2024. "*Joking, prophetically*: The use of humour in Mormon leaders' sermons." *Comedy Studies* 15:77–93. https://doi.org/10.1080/2040610X.2023.2263706.

Harrington, Jonathan, Sallyanne Palethorpe, and Catherine I. Watson. 2000. "Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts." *Journal of the International Phonetic Association* 30:63–78. https://doi.org/https://doi.org/10.1017/S0025100300006666.

Knowlton, David. 1991. "Belief, metaphor, and rhetoric: The Mormon practice of testimony bearing." *Sunstone* 15 (1): 20–27.

Madsen, Grant, and Joseph Corey. 2023. "What's in a name?: The growing focus on Jesus Christ (by name) since 2000 in general conference talks." *BYU Studies Quarterly* 62 (4): 141–153.

Pichler, Heike, Suzanne Evans Wagner, and Ashley Hesson. 2018. "Old-age language variation and change: Confronting variationist ageism." *Language and Linguistics Compass* 12:e12281. https://doi.org/https://doi.org/10.1111/lnc3.12281.

Sankoff, Gillian. 2018. "Language change across the lifespan." *Annual Review of Linguistics,* 297–316. https://doi.org/https://doi.org/10.1146/annurev-linguistics-011817-045438.

———. 2019. "Language change across the lifespan: Three trajectory types." *Language,* 197–229. https://doi.org/10.1353/lan.2019.0029.

Sankoff, Gillian, Hélène Blondeau, and Anne Charity. 2001. "Individual roles in a real-time change: Montreal (r->R) 1947–1995." *Etudes & travaux,* 141–157.

Teerlink, Mabel. 2024. "Covenant communication: A corpus study of shifts in Latter-day Saint discourse." *Schwa* 30:1–9.

Trudgill, Peter. 1983. "Acts of conflicting identity: The sociolinguistics of British pop-song pronunciation." In *On dialect: Social and geographical perspectives,* edited by Peter Trudgill, 141–160. New York, New York: New York University Press.

van de Velde, Hans, Marinel Gerritsen, and Roeland van Hout. 1996. "The devoicing of fricatives in Standard Dutch: A real-time study based on radio recordings." *Language Variation and Change* 8:149–175.

Wenger, Etienne. 1998. *Communities of practice: Learning, meaning, and identity.* Cambridge, England: Cambridge University Press.

Williams, Peter. 2024. "Detecting semantic differences between LDS and Christian speech." *Schwa* 30:61–73.

Yaeger-Dror, Malcah. 1993. "Linguistic analysis of dialect 'correction' and its interaction with cognitive salience." *Language Variation and Change* 5:189–224.