Schwa

Language and Linguistics

Cover design by Halli Romero.

Contents

Staff

Editor in Chief
Rachel Hart

Managing Editor
Shaylyn Lawyer

Design Director
Abby Ellis

Senior Editors
Halli Romero
Alyssa Brownigg
Abby Ellis

Staff Editors
Taylor Wood
Ciera Griggs
Elise Green

Design Team
Rachel Hart
Shaylyn Lawyer
Taylor Wood
Halli Romero
Elise Green

Faculty Advisor
Dr. Dirk Elzinga

About *Schwa*

We are an academic journal produced by the students of Brigham Young University. Our mission is to increase the amount and accessibility of linguistic scholarship—especially for those without graduate school experience—while simultaneously training editors and designers in the ways of modern publishing. Some of our articles are strictly theoretical and academic. Others are less technical and more personal in nature. Experiments, surveys, corpus analyses, and essays are all acceptable. We have published on all the following subdisciplines of linguistics and more:

- Phonetics, the perception and production of speech sounds
- Phonology, the system of speech sounds used in a given context
- Semantics, the meaning constructs of words and sentences
- Syntax, the structure of permissible and meaningful sentences
- Pragmatics, real-world language use and other speech-related actions
- Sociolinguistics, language variation based on sociological factors
- Psycholinguistics, the cognitive tasks necessary for language
- Fieldwork notes from living in a foreign language-speaking community
- Forensics linguistics, the role of language in law

We are always accepting submissions. Articles on any language are welcome, including cross-linguistic studies, but they must be written in English.

Our staff includes both editors and graphic designers. We extend an open invitation for new staff members. Go to our website at schwa.byu.edu to submit an article or join our staff.

Editor's Note

When I walked into the student journal fair almost two years ago, I would never have guessed that I would be writing this editor's note just a few semesters later. I was a shiny, brand-new editor, looking to gain experience and maybe learn a thing or two about linguistics along the way. But after four semesters on the *Schwa* staff, I think I can safely say that I've learned more than just a thing or two.

Truth be told, this hasn't been an easy note to write. There have been many false starts and bouts of writer's block. How could I possibly sum up four semesters of working with amazing editors and authors, learning, and growing? I can't; I can only thank those who have made it possible.

First of all, I am grateful for all our editors and the work they have done to bring this issue of *Schwa* to life. Each one has dedicated so much of their time and talents, and for that, I thank them. I'd like to thank Shaylyn, my amazing managing editor, as well as Abby, the previous editor in chief, who has spent so much of her time this semester helping me grow into this new role.

As always, we are grateful for our wonderful authors, who have shown courage and vulnerability in putting their works out there for so many to see. Without them, we would have no journal to publish.

We are also grateful for the Linguistics department and our professors. We thank them for the hard work they put in every day, helping us learn and grow. We are especially grateful to our faculty advisor, Dr. Dirk Elzinga, for his guidance and counsel.

And of course, we'd like to thank you for taking the time to read issue 30 of *Schwa: Language and Linguistics*. We hope you enjoy!

Rachel Hart
Editor in Chief

Covenant Communication

A Corpus Study of Shifts in Latter-day Saint Discourse

Mabel Teerlink

This article investigates historical changes in discourse about covenants among leaders of the Church of Jesus Christ of Latter-day Saints. By analyzing both the prevalence and the collocates of the term covenant *in a corpus of general conference talks from the 1850s through the 2020s, this article finds that the word's relative frequency and contextual usage have varied significantly over time. These findings reflect the ongoing development of religious discourse and the power of corpus linguistic tools for revealing discourse patterns.*

The concept of covenants has long been important in Judeo-Christian tradition, from the Old Testament account of Abraham's agreement with God, to the New Testament account of Jesus's "blood of the covenant" (Mark 14:24 NIV). Like other Christian sects, the Church of Jesus Christ of Latter-day Saints has taught about covenants since its inception—one of the Church's main books of scripture is even called The Doctrine and Covenants—but in recent years, the Church has seemed to place special emphasis on these teachings. For example, talks in the Church's October 2023 general conference included expressions like "covenant path," "covenant people," "covenant relationship," and "covenant blessings." This study examines expressions such as these in a corpus of general conference talks to analyze how discourse about covenants has changed over time.

Literature Review

Previous studies of linguistic changes in Latter-day Saint discourse reveal noteworthy patterns. Tenney (2017) finds that over time, both gendered references and speaking opportunities in general conference have gradually shifted from favoring men toward more equality between the genders. Hilton and Barringer (2019) analyze the general conference discourse on the topic of Jesus Christ's suffering in the Garden of Gethsemane. They find that references to Gethsemane have increased over time and that the themes surrounding Gethsemane have changed over time, although they draw these conclusions from raw frequency counts rather than normalized data. Stapley (2021) uses normalized frequencies to analyze how the endings of general conference talks have fluctuated over time and finds that invoking the name of the Lord is a development that began in the 1950s and continues to the present.

While regular general conference listeners may have sensed a recent increase in talks about covenants, no research has yet attempted to quantify this increase or analyze it in depth. This study attempts to fill that gap by investigating how the normalized frequencies of words related to and surrounding covenants have changed over time.

Data and Methods

This study uses Mark Davies's general conference corpus at lds-general-conference.org, which includes all recorded general conference talks from the 1850s onward. The corpus also has convenient features for comparing time periods and finding collocates, or words that occur near the word in focus.

The search term "covenant*" was used as a lemma, or collection of word forms based on the word *covenant*. The terms resulting from this search are shown in table 1.

Table 1

All forms of the covenant *lemma included in the corpus search*

covenant
covenants
covenanted
covenant-keeping
covenant-making
covenantal
covenanting
covenant-breakers
covenant-breaking
covenant-breaker
covenanters
covenantors
covenanteth
covenanter
covenant-taking
covenant-makers
covenant-maker
covenant-keepers
covenant-consistent

To analyze changes in overall use of the lemma, the relative frequency (occurrences per million words) of all the above terms combined was recorded and plotted for each decade from the 1850s through the 2020s and for each individual year from 1852 to 2023.

To analyze changes in words surrounding the lemma, Davies's collocates feature was used to find the most frequent collocates (within four words to the left and four words to the right of the *covenant* lemma) for each decade from the 1850s to the 2020s. This feature assigns each collocate a mutual information (MI) score of significance, which is calculated using the frequency of the node words, the frequency of the collocate, and the size of the corpus. After some experimentation, the minimum MI threshold for this study was set to 2, which excluded articles, prepositions, and some verbs. The ten most frequent collocates for each decade (with a minimum MI score of 2) were then collected.

Results

Figure 1 shows the relative frequency of the *covenant* lemma for each decade from the 1850s through the 2020s. Figure 2 shows a more fine-grained record of the relative frequency for each year from 1852 to 2023. With some variation, the frequency stays fairly low until a small spike in the 1950s, followed by a greater spike in the 1990s, followed by a large increase in the 2010s, and the greatest increase in the 2020s. The 2020s only include data from 2020 to 2023, so by the end of this decade, the relative frequency could be even greater. Generally, it seems that the *covenant* lemma has been increasing among conference talks from the 1950s onward.

Tables 2 and 3 show the ten most frequent collocates of the *covenant* lemma (with a minimum MI score of 2) for each decade from the 1850s through the 2020s. One unexpected influence on the data is the frequency with which general conference speakers use the word *covenants* when citing the Doctrine and Covenants, one of the Church's books of scripture. *Doctrine* is among the three most frequent collocates in every decade, and *book* was a similarly frequent collocate up until the early 1900s, presumably from usages like this:

> "Here are the Bible, the Book of Mormon, and Book of Doctrine and Covenants." –Heber C. Kimball, 1861

Figure 1

Relative frequency of the covenant *lemma by decade*
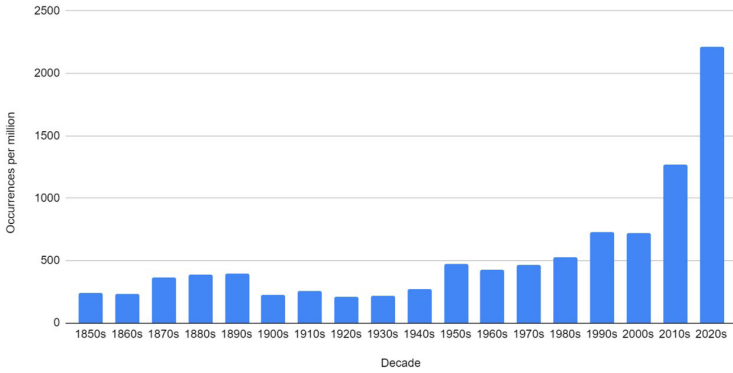
Figure 2

Relative frequency of the covenant *lemma by year*
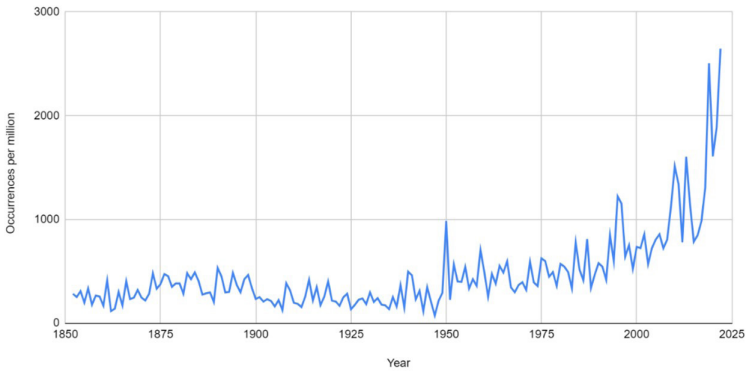
Table 2

Top ten collocates of the covenant *lemma by decade, 1850s–1930s*

Rank	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s
1	book	doctrine	doctrine	doctrine	doctrine	made	made	doctrine	doctrine
2	doctrine	book	book	book	book	doctrine	everlasting	made	everlasting
3	made	made	made	everlasting	made	book	doctrine	everlasting	made
4	everlasting	everlasting	everlasting	new	everlasting	everlasting	new	new	entered
5	new	new	new	made	entered	entered	keep	entered	marriage
6	make	keep	entered	entered	new	new	league	broken	new
7	keep	faithful	marriage	marriage	oath	keep	born	hath	enter
8	broken	broken	enter	abide	faithful	make	marriage	mine	book
9	Abraham	sec.	keep	keep	make	sacred	book	marriage	broken
10	fulfil	marriage	broken	mediator	keep	commandments	entered	oath	Abraham

Table 3

Top ten collocates of the covenant *lemma by decade, 1940s–2020s*

Rank	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s	2020s
1	everlasting	doctrine	doctrine	doctrine	doctrine	doctrine	doctrine	doctrine	path
2	made	made	everlasting	everlasting	made	made	made	keep	doctrine
3	doctrine	everlasting	made	made	ordinances	sacred	keep	make	ordinances
4	marriage	marriage	keep	oath	oath	keep	sacred	made	make
5	new	new	marriage	marriage	keep	ordinances	ordinances	sacred	keep
6	keep	keep	oath	new	make	make	temple	path	sacred
7	oath	make	sacred	keep	marriage	keeping	make	ordinances	keeping
8	entered	obligations	make	belongeth	everlasting	marriage	keeping	keeping	made
9	broken	oath	new	make	sacred	temple	oath	temple	temple
10	sacred	section	section	pearl	new	new	renew	making	stay

However, *book* decreased in frequency after the 1910s, perhaps because general conference speakers began to simply refer to the Doctrine and Covenants without *book* as part of its title. The collocate *section* or the abbreviation *sec.* also appears in the top ten for the 1860s, 1950s, and 1960s because of references to sections of the Doctrine and Covenants.

It should be noted that mentions of the Doctrine and Covenants are included in the overall frequency of the *covenant* lemma and therefore may affect frequency trends. For instance, a rise in overall frequency may be because of increased citations of the Doctrine and Covenants rather than increased references to covenants in general. Separating Doctrine and Covenants references from the other data could help distinguish how much these citations affect the overall frequency. However, this would require more advanced searches than are possible through the Davies corpus, so for the purposes of this study, the references are included as part of the data.

The phrase "new and everlasting covenant" appears to be one of the most common uses of *covenant* through the 1980s, but it was preempted by other collocates from the 1990s onward. Similarly, the collocate *marriage* is in the top ten for most of the nineteenth and twentieth centuries but does not appear in the top ten from the 2000s onward.

The verbs *enter, make,* and *keep* are high-ranking collocates, but *enter* is not among the top ten from the 1950s onward, whereas *make* and *keep* are.

Broken and *oath* were among the top ten collocates for several decades but then declined. Other collocates, such as *sacred, ordinances,* and *temple,* first appeared in the top ten during the second

half of the twentieth century and remained prominent until the 2020s.

Some collocates only appear in the top ten for one decade, such as *fulfil* in the 1850s, *abide* and *mediator* in the 1880s, *commandments* in the 1900s, *league* and *born* in the 1910s, *hath* and *mine* in the 1920s, *obligations* in the 1950s, *renew* in the 2000s, and *stay* in the 2020s.

Notably, *path* first appears as a top-ten collocate in the 2010s and is now the most frequent collocate in the current decade. This also explains the rank of the collocate *stay* in the 2020s; an increasingly frequent phrase is one such as this:

"We endure to the end and stay on the covenant path." –M. Russell Ballard, October 2021

The first instance of the phrase "covenant path" in the corpus is found in a 2007 conference talk that quotes a line from a 2006 Church magazine article by Jeffrey R. Holland:

"The promptings of the Holy Ghost will always be sufficient for our needs if we keep to the covenant path."

The phrase was used next by D. Todd Christofferson in 2009:

"Early Church leaders in this dispensation confirmed that adhering to the covenant path provides the reassurance we need in times of trial."

The phrase then rose in frequency throughout the 2010s and 2020s and will presumably continue to increase.

Overall, it appears that phrases like "new and everlasting covenant" and collocates like *broken, oath,* and *marriage* have generally decreased over time, whereas phrases like "covenant path" and collocates like *sacred, ordinances,* and *temple* have generally increased over time. Also, the actions that are applied to covenants seem to have transitioned away from verbs such as *enter* and *break* and shifted toward verbs such as *make* and *keep.*

These changes perhaps follow a pattern of a decreased emphasis on covenants as one-time contracts with God and an increased emphasis on covenant keeping as an ongoing way of life. The shift could also reflect a growing inclusivity in the Church, as marriage is not necessarily a reality for everyone, but participation in proxy temple ordinances is available to all Church members.

Discussion and Conclusions

This article adds to the small body of linguistic research on the diachronic development of Latter-day Saint discourse. Like previous studies, this study has shown that linguistic patterns in general conference talks have changed over time. Specifically, this study's findings indicate that the relative frequency of words related to covenants has generally increased over time, and that the collocates surrounding those words have varied over time in frequency and focus.

These findings may help members of the Church of Jesus Christ of Latter-day Saints (and those who interact with them) to be more aware of ongoing developments in Church teachings and communication. The Church professes to be a "true and living" organization, and this study and others show that its general conference discourse is indeed alive and actively evolving. Further research may reveal other linguistic shifts in Latter-day Saint discourse or possible causes for those shifts. Much remains to be explored in this area and other areas of religious discourse through the lens of corpus linguistics.

References

Ballard, M. Russell. (2021). Lovest thou me more than these? *October 2021 General Conference*. https://www.churchofjeuschrist.org/study/general-conference/2021/10/31ballard?lang=eng.

Christofferson, D. Todd. (2009). The power of covenants. *April 2009 General Conference*. https://www.churchofjesuschrist.org/study/general-conference/2009/04/the-power-of-covenants?lang=eng.

Hilton, John, III & Joshua P. Barringer. (2019). The use of Gethsemane by church leaders, 1859–2018. *BYU Studies Quarterly. 58*(4), 49–76. https://scholarsarchive.byu.edu/facpub/3671/.

Holland, Jeffrey R. (2006, October). What I wish every new member knew—And every longtime member remembered. *Ensign*. https://www.churchofjesuschrist.org/study/liahona/2006/10/what-i-wish-every-new-member-knew-and-every-longtime-member-remembered?lang=eng.

Kimball, Heber C. (1861). Cultivation of the spirit of truth—Trials of the church, &c. *Journal of Discourses. 9*(11). https://journalofdiscourses.com/9/11.

Nelson, Russell M. (2019). Spiritual treasures. *October 2019 General Conference*. https://www.churchofjesuschrist.org/study/general-conference/2019/10/36nelson?lang=eng.

Palayon, Raymund T., Richard Watson Todd, & Sompatu Vungthong. (2021). A corpus-based study of sermons to determine the stage of religious development. *Proceedings of the International Conference: DRAL 4,* 73–93. https://sola.pr.kmutt.ac.th/dral2021/wp-content/uploads/2022/06/6.pdf.

Stapley, Jonathan. (2021). Invoking the name of the Lord: A quantitative study. *Journal of Mormon History, 47*(4), 1–21. https://doi.org/10.5406/jmormhist.47.4.0001.

Tenney, Devon. (2017). Speaking of gender, generally: Analysis of gendered references and speaking opportunities in LDS general conferences. *FHSS Mentored Research Conference*, 300. https://scholarsarchive.byu.edu/fhssconference_studentpub/300.

Clipping in French and Japanese

Megan Hamilton

Contemporary linguistics largely ignores the sociolinguistic factors behind clipping. However, especially when considering languages like French and Japanese that frequently clip, questions of who is behind these changes arise. Current research shows that younger speakers are mainly responsible and that multilingualism also has a strong effect on language change, especially in Japanese and French clippings. Computer-mediated communication (CMC) has also created widespread, rapid language change. All of these sociolinguistic factors are likely sources of clipping. Interviews with younger speakers and corpora built from CMC with native Japanese and French speakers could render it possible to determine which sociolinguistic factors are at work.

Clipping, or truncation (the process by which speakers of a language shorten a word, such as *refrigerator* becoming *fridge* in English) as a whole has been extensively studied from a morphological standpoint in the linguistics community, with complex talk of collocates, aphesis, and stressed and unstressed vowels. Yet there's almost no discussion to be found on the sociolinguistic factors involved with clipping, no matter the language in question. There is some mention of speech styles and social identity as driving factors, but it remains quite limited. When looking at the literature on other types of word formation processes, there is a lot of research available on age, gender, multilingualism, and other factors contributing to the creation or repurposing of words in a given language. It seems likely that such factors contributing to one form of word formation would leak into others that are closely related, such as clipping. This article will look at French and Japanese, languages that seem to be prime territory for clipping, and develop research on the demographics of clipping.

There is ample literature that backs up the claim that French and Japanese clip more than other languages—this kind of research shows that young French and Japanese speakers truncate their words more than young speakers of other languages, and other literature details the tendency Japanese has to clip borrowed words. It seems that perhaps multilingualism could be a major factor in the *who* of such clippings, as adolescents who are able to speak more than one language are often more innovative with their language use than others. Their use of L1 (native language) grammar in L2 (second language) discourse could also be a potential factor of clipping in Japanese and French since both languages clip borrowed words quite frequently. Some frequent clips in French are *appartement* (apartment) becoming *appart,* or *exactement* (exactly) becoming *exact,* and common Japanese clips are コンビニエンスストア (*konbiniensusutoa,* convenience store) becoming コンビニ (*konbini*), or 東京大学 (*toukyo daigaku,* Tokyo University) becoming 東大 (*toudai*). No matter the cause, based on similar research, it seems most likely that the two main factors that determine who will use clippings are age and multilingualism. Multilingualism brings in grammar and pronunciation factors that are being transposed onto these clippings, and younger age brings more language innovation. Yet these innovators are not always seen in the best light, as responses to teen language innovation as seen in computer-mediated communication studies have been

less than warm. All of this shows that there is an abundance of research on the who and why of similar language situations, but barely any on the who and why of clipping—a gap this article hopes to fill.

An Overview of Clipping

Clipping, according to Don, is largely unpredictable in terms of how it will break down the morphology of any given word (2014). It can affect a word in several different ways, be it back clipping (*metropolitan* becoming *metro*), front clipping (*telephone* becoming *phone*), or mid clipping (*refrigerator* becoming *fridge*). The register of many clippings appears to remain fairly informal and closely related to the speech of adolescents, particularly slang speech. Hilpert et al. (2021) list some example clippings like *sis*, *bro*, *shrooms*, and *legit*; although one of their other examples includes *fridge*, which is ubiquitous in general American English. Within a population of clippings taken from Wikipedia as a part of Jamet's 2009 study, ninety percent of them belong to the nominal group, and seventy-five percent of them experience end-clipping (also known as back clipping) (Hilpert et al., 2021). So, contrary to what Don may have thought, it does seem that there is some amount of predictability surrounding clipping, though it does not always stick within these boundaries.

Hilpert et al. mention shortly after this that clipped forms often refer to highly specific referents which may serve as a way to distinguish one's in-group through language. Similarly, clipping has been used with personal names as a way of "individuating" those people in question (Fajardo, 2018). Fajardo suggested several global factors motivating clipping of personal names in his study, namely economical speech, stylistic choice, and expressiveness, among others. So though there is not much literature on the sociolinguistic forces behind clipping, what is there seems to indicate it may have to do with speech style and identity within one's speech community. That said, there still is not much hard evidence or empirical research on the topic.

When looking at the morphology of clipping cross-linguistically, Arndt-Lappe claims that there are "formal properties of truncation that seem to hold across languages, to the effect that truncatory patterns, even in typologically distinct languages, look strikingly similar" (2018, p. 147). Arndt-Lappe then qualifies his statement, adding that within these universal constraints, languages will

make use of clipping based on how their own distinct grammar works. In other words, every language employs the same tools, but with slight variation in the results—very similar to the conclusions made by Hilpert et al. Once again, however, other researchers came in with evidence to the contrary. When Demuth and Johnson (2003) studied clipping in young children's speech, there did not appear to be as many cross-linguistic constants. In fact, it appeared that the level of clipping and types of clips depended largely on the language that was being spoken. It was found that syllable density and articulation difficulty were predictable, major factors behind clipping in young children, but that these factors still had great variation depending on the language.

Up until this point, research on clipping has been morphological, not sociolinguistic, even on the few threads of social identity that were mentioned by Hilpert et al. and Fajardo. That does not indicate that there is no trace of sociolinguistic factors, nor what kind of research would target demographic studies on clipping.

Multilingualism

Picking up where Demuth and Johnson left off, it does seem that one of the biggest factors on clipping and general word formation is multilingualism. According to Darquennes et al. (2019), there are several factors playing into language change driven by multilingualism. The demographic most affected by and perpetuating word formation is those younger speakers who have yet to attain puberty. They have "more evidence for both retention and substitution of L1 case assignments than in adults—more evidence for retention and substitution than for loss" (Darquennes et al., 2019, p. 175). Bilingualism of course leads to code-switching; however, what is perhaps more interesting is the transplant of L1 features onto the child's L2. Though code-switching can lead to the introduction of new words into a language via borrowing, it is this mapping of L1 grammar to a child's L2 that could be a factor in clipping.

Irwin treated a similar subject when talking about mora clipping in Japanese. Unlike English, Japanese words are built up of little consonant-vowel pairs called moras that are given equal length in pronunciation. It seems that there is a strong tendency to clip when it comes to borrowed words in Japanese, yet it is typically the younger generation that is employing those words to begin with (2011). There tend to be several different clippings for

any given borrowed word, (often from English) but one clipping tends to be favored over the others. Irwin used the Romanized *konbiniensusutoa* (convenience store) as an example of this concept, with the favored clipping being *konbini*, but with another (less successful) variant, *sutoa*, coexisting in modern Japanese. This seems like it could potentially be mirroring Japanese's ready acceptance of multiple readings for one character in its own language. However, though some of the readings are tied to specific vocabulary words and structures, there are still coexisting readings that are preferred over others. So, it seems that these changes are a possible consequence of younger speakers' innovations as they learn languages such as English and begin mapping L1 features to their L2.

Another observation made by Darquennes et al. was that among adolescents who are in a period of identity searching, there is a high flexibility of language norms and high tolerance for linguistic variation (2019). And it is during this period that adolescents who have access to multiple languages will be most likely to swap features in and out of their languages, mainly pulling from their L1 to innovate in L2. Bilingualism in children leads to more innovation in language, and it is possible that this innovation could be contributing to the sociolinguistic motivations behind clipping.

Darquennes et al. also discussed multiethnolects, urban youth speech styles that are the result of the mixed speech of speakers with different heritage languages (2019). These multiethnolects include "non-standard and innovative phonetic, grammatical, and discourse-pragmatic features" (Darquennes, 2019, p. 177). Thus, it is a logical step to assume that such multilingual speakers of French and Japanese would be more prone to clipping borrowed or native words as a part of mixing L1 and L2 features.

Truncation and Age

Young Learners

Moving to age as a factor, there has been extensive research into clipping in both young French speakers and young Japanese speakers, as it has been found that both languages' complex articulation stumps younger speakers. Ota (2006) specifically used Japanese as the language of research because of the large amount of high syllabic words contained in the language, assuming that, more than anything, this would be the main determining factor

as to whether or not a child would truncate a word in their language. Demuth and Johnson pointed out, however, that if this is simply a question of word length or sound restriction, "we expect these maturational limitations to be found cross linguistically" (2003, p. 212). Demuth and Johnson follow this comment by citing research to the contrary which shows that children learning Spanish as their first language do not share the articulation difficulties that the French children exhibited. Ota found that though some of the patterns found in French child truncation were uncommon in Japanese, the more syllables a Japanese word had, the more likely a child would be to truncate it (2006).

Childhood difficulty with articulation cannot entirely be considered conscious or systematic clipping, as most of the clippings produced by children are eventually removed as they mature. However, it does give some indication that the more syllables a word has, the more likely a speaker is to truncate it in some way. Given that Japanese and French are known to be syllable dense and full of long words, it is no surprise that both of these languages would see a fair amount of truncation in early childhood development. It is also possible then that older children and adolescents would retain some of these truncations, though it is a subject lacking in research. Teen language innovation as a general topic, on the other hand, has seen a lot of research.

Teens

Teens are often seen as the "prime innovators of non-standard speech form" and as such have received a lot of attention with research on language contact, linguistic innovation, and change (Darquennes et al., 2019, p. 175; Cutler & Røyneland, 2018). It is evident why teens, who have a high tolerance to linguistic variation, are the demographic with a strong likelihood to innovate as a part of their identity searching. Darquennes et al. also claim that younger children and young adults are those most likely to swap new features in and out of a new language as they develop their social identity (specifically referring to multilingualism) as opposed to adults (2019).

In modern language innovations in teens, it is impossible to avoid the topic of computer-mediated communication (CMC), as this is the homeland of major modern language changes. Such innovations can be seen as well in instant messaging styles of communication employed by teens (Tagliamonte & Denis, 2008).

Tagliamonte and Denis note that the innovations in language use seen in this CMC are mainly developed and spread by teens (2008). Unsurprisingly, the majority of innovations made via CMC are ones that relate to spelling. Spelling innovations that began as a way to text message despite character limitations and tedious clicking of buttons have contributed to modern language use online, which has only continued to distinguish itself from spoken speech or more formal registers of writing. Some examples of such distinctions that Tagliamonte and Denis collected in their 2008 research include "lol y dont i believe u" or "i dont kno nothing really" (p. 3). Though Tagliamonte and Denis do not explicitly bring in the concept of clipping, this written reduction of words by teens is just that. In this case, it is inaudible, and it is not leading to the creation of new words to the lexicon, but it is undeniably a new form of clipping used mostly by teens. Other researchers such as López Rúa argue that all forms of word shortening such as abbreviations, initialisms, and blends could be grouped together with clipping in a more general description of such word formations (2002).

It is clear that through CMC, teens are the biggest innovators of language and that these new additions to their language are not favored by parents and teachers alike. Not only are they often a clear indicator of age, but they are also not seen in a favored light by those out of the in-group that uses them. This could potentially mean that clippings introduced and perpetuated by teens are not valued by an older demographic, and thus would be unique to teen speech—another potential clue as to who the main speakers using clippings are. This new domain of computer-mediated language innovation remains almost exclusive to teens, but once more there's no specific treatment on clipping (or similar processes) and whether it also remains mostly used by teens as opposed to other age demographics.

Future Research

As can be seen through the literature discussed thus far, there is a large gap in research when it comes to sociolinguistic studies and clipping. Though there has been observation of the relation between in-group identifiers and clipping, it is more of a brief overview of the relationship between those two. There has been very little research determining a link between clipping and age. Especially when it comes to languages like French and Japanese,

for which there has been some research on the types of clipping and their unique tendencies with clipping, there is a lot of room for research. My hypothesis is that multilingual adolescents are the main group using clippings, and are also going to be the group that has the most new clippings.

To test this, my eventual goal would be to do a comparative study between clipping in French and Japanese since there is literature confirming that they share some similar features. Logistically speaking, however, I think it would make the most sense to start by doing this research in Japanese as a sort of pilot to see how age and multilingualism plays into clipping, and then after getting those results, if it is successful, shift to doing the same in French. Given that there is already some literature showing ties with mora clipping and borrowed words in Japanese, I think it is a stronger initial thread to start with.

Ideally, I would work with people of all ages twelve and up, with age groups split into blocks of fifteen to twenty years. Seeing as it seems that a lot of newer clippings come from written CMC, I think it would be best to start this research via survey responses, collecting surveys that represent each age group as equally as possible which would allow me to determine which age group is more likely to clip words. I would make sure that demographic elements such as gender and multilingualism were tracked and logged in order to get the clearest picture of who is developing these clippings. Occupation would also be tracked due to the suggested in-group nature of clipping, but it would not be the primary target of the survey. In the survey I would include questions that ask to elicit specific words; for example, "How would you refer to a small store in your neighborhood" as a way to elicit the Japanese *konbini*. After that, I would ask survey respondents to explain their hobbies, once as if they were talking to a future employer, and once as if they were talking with a friend. The goal there would be to elicit in-group clippings and determine how ubiquitously they are being used.

Conclusion

It is clear that there is a plethora of research on age and language use, language innovation and multilingualism, and the morphology of clipping, but there is very little research connecting the three. What we do know is that similar processes are frequently used by teens in computer-mediated communication and

that these innovations are not seen in a favorable light by older generations. In addition, there is also lots of research on how multilingualism affects language innovation, and how the age of multilinguals plays into this as well. All of these treated topics ignore clipping, but they are a good place to start the groundwork. I think that the best place to start with this new research is by studying French and Japanese, as they seem to be breeding grounds for clipping. Via a survey approach, it would be possible to get a lot of data quickly and determine sociolinguistic factors at play.

References

Arndt-Lappe, Sabine. (2018). *Expanding the lexicon: Linguistic innovation, morphological productivity, and ludicity.* De Gruyter. https://doi.org/10.1515/9783110501933-143

Cutler, Cecelia, & Unn Røyneland. (2018). *Multilingual youth practices in computer mediated communication*. Cambridge University Press. https://doi.org/10.1017/9781316135570

Darquennes, Jeroen, Joseph C. Salmons, & Wim Vandenbussche. (2019). *Language contact: An international handbook.* De Gruyter Mouton. https://doi.org/10.1515/9783110435351

Demuth, Katherine, & Mark Johnson. (2003). Truncation to subliminal words in early French. *Canadian Journal of Linguistics/Revue canadienne de linguistique, 48*(3–4), 211–241. https://doi.org/10.1017/S0008413100000657

Don, Jan. (2014). *Morphological theory and the morphology of English.* Edinburgh University Press. https://ebookcentral.proquest.com/lib/byu/detail.action?docID=1717552

Fajardo, José Antonio Sánchez, & Félix Rodríguez González. (2018). Motivations and morphological variations in clipped personal names: A crosslinguistic approach. *Lingua, 206*, 35–48. https://doi.org/10.1016/j.lingua.2018.01.006

Hilpert, Martin, David Correia Saavedra, & Jennifer Rains. (2021). A multivariate approach to English clippings. *Glossa, 6*(1), 1–30. https://doi.org/10.16995/glossa.5771

Irwin, Mark. (2011). Mora clipping of loanwords in Japanese. *Journal of Japanese Linguistics, 27*(1), 71–81. https://doi.org/10.1515/jjl-2011-0105

López Rúa, Paula. (2002). On the structure of acronyms and neighbouring categories: A prototype-based account. *English Language and Linguistics, 6*(1), 31–60. http://doi.org/10.1017/s136067430200103x

Ota, Mitsuhiko. (2006). Input frequency and word truncation in child Japanese: Structural and lexical effects. *Language and Speech, 49*(2), 261–295. https://doi.org/10.1177/00238309060490020601

Tagliamonte, Sali A., & Derek Denis. (2008). Linguistic ruin? Lol!
    Instant messaging and teen language. *American Speech, 83*(1),
    3–34. https://doi.org/10.1215/00031283-2008-001

In My Linguistics Era

A Study of Era Usage in the Context of Taylor Swift

Alayna Beck

Taylor Swift's Eras tour increased the word era*'s frequency of use and influenced its definition. This article examines this semantic shift through corpus and survey research methods. The author concludes that the new definition of* era *has not overtaken the original, though there is potential for it to become more widespread. As of 2023,* era *has not become a slang term, but again, it has the potential to become so over time as the Eras tour and the use of the word continues.*

I n Taylor Swift's eighth studio album, she sings, "You taught me a secret language I can't speak with anyone else" ("illicit affairs," 2020). Taylor Swift may have had a metaphorical secret language with a previous lover, but she has also created a not-so-secret language with the whole world. This pop icon has certainly had her moment this year. Her music and recent Eras tour have changed many aspects of the world, including boosting America's economy (Kopstein, 2023). The tour was an amalgamation of all ten of her albums—each album representing an era in her career. Since Swift's announcement of the tour in 2022 (Ray, 2023), there has been an observable shift in the usage of the word *era* in the general public.

This article will seek to understand the shift in the definition of *era* in colloquial language since Swift's tour announcement. Gaining this knowledge will provide insight into the impact that pop stars on Taylor Swift's level have on language. Awareness of important influences on the English language is crucial to understanding future language shifts. I hypothesize that *era* has begun to shift to a slang term, but it is not widespread and has not yet replaced the original meaning of the word.

Literature Review

The phenomenon of Taylor Swift has been studied through many linguistic lenses. Swift has had her own linguistic changes throughout her career, beginning in 2014 when she released her single "Tim McGraw" (Ray, 2023). Linguists remark on her code-switching in her voice and lyrics to match different genres and reach different audiences over the progression of her career. Her noticeable country accent in earlier albums has disappeared in recent years to appeal to those who now listen to her music, a type of style shifting calling attention to the audience (Luu, 2020). Ignoring the genre change, her lyrics have also been extensively studied for their strength in various figurative language methods (Siagallan, 2017). Swift has a long linguistic history and will no doubt go on to make more language-defining moments.

One such recent moment is her Eras tour, which began in March of 2023 and is continuing internationally at the moment (Ray, 2023). The word *era* originated in the mid-seventeenth century and has maintained a consistent definition through time. It has always meant an extended period marked by a distinctive feature (Oxford, 2023). That was the case until Taylor

Swift announced her Eras tour. Taylor Swift fans (Swifties) adopted the word as a slang term to mean a shorter time in one's personal life (Cavender, 2022). You may hear "I'm in my procrastination era," or "Welcome to my corporate era," most commonly on TikTok and perhaps Instagram. As internet language tends to do, *era* used in the slang sense has crept into the colloquial language of Gen Z-ers. This study will further examine the extent to which the definition of *era* has changed. In studying the definition change of *era*, these results will aim to cover what other linguists have yet to do regarding Taylor Swift.

Methods

To investigate the hypothesis, I conducted two methods of research. The starting method was a survey. The first block of questions contained typical demographic questions to get a background on the participants, including inquiries about age and gender. In addition, participants were asked if they considered themselves a Swiftie to gain their general opinion of the artist. On the next page, participants were informed that Taylor Swift announced her Eras Tour on November 1, 2022, so they would have an accurate timeline reference for the following questions, which were: "How often did you use *era* before Swift announced the Eras Tour?" and "How often do you use *era* after Swift released the Eras Tour announcement?" Participants were given sliders ranging from 0–100 to rank their frequency of use. After, in order to get straight to the point, the survey directly asked whether or not they considered *era* to be a slang term.

The next block of questions addressed the acceptability of different usages of *era*. Sentences from the News on the Web (NOW) corpus were used for examples. This corpus is updated almost daily with words from online sources, such as blogs, websites, and social media. The first sentence chosen was pre-Eras tour while the second sentence was post-Eras Tour. For both, the survey asked: "How acceptable is this sentence?" and "Could you see yourself saying this sentence?" Participants were given the same 0–100 sliders for these questions.

For the second method of research, I returned to the NOW corpora and conducted a simple chart search: [my * era] to match the observation of the new usage of *era*. This search provides a look at how the definition of *era* has changed in colloquial language from

a distinct, extended time in history definition to a short-term, personal period definition, synonymous with phase or stage.

Results

First, here are the results of the NOW corpus search. The increased use of *era* in its new, evolving definition is clear. [My * era] remained at one percent of one million words from 2016 through 2021 until 2022, the year Taylor Swift announced the Eras tour, where it increased to three percent. In 2023, the number of words per million doubled, resulting in six percent of one million words (see figure 1).

While these results are very telling of the concrete increase in *era* usage, this method of research does not reflect the sociolinguistics of this definition shift. The survey, this study's second research procedure, demonstrates the general public's belief in regards to how they use *era* and how they feel about the shifting usage. Out of fifty-six responses, the majority of participants fell between the ages of sixteen and thirty, with seventy-three percent identifying as female. Most surveyees, seventy-seven percent to be exact, did not claim to be a Swiftie. This non-Swiftie demographic may have a bias, affecting numbers and possibly leading to differences in results if the study is duplicated.

Figure 1

*[My * era] usage through the years*
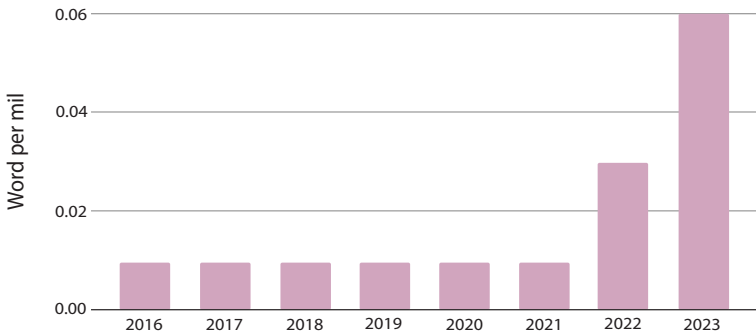
After establishing the demographic background, the following block of questions inquired about the participants' usage of era and thoughts about the word being considered slang. The average ranking on a scale from 0 to 100 of using *era* in casual conversation before Taylor Swift announced her tour was 14.19. The reported number after her tour announcement was 23.66 (see figure 2). That is a 66.7 percent increase. When asked about *era*'s status as a slang word, sixty percent of participants did not consider *era* to be a slang word.

Figure 2

Era *usage before and after the Eras tour*

Ranked frequency on a scale from 0–100
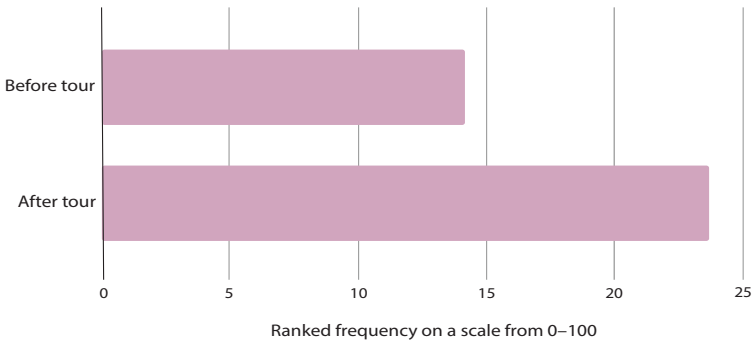
Then respondents answered acceptability questions and reported how likely they were to use certain sentences, either spoken or written. The first sentence taken from the NOW corpus was "A whole second era of aviation was opened up by Whittle and yon Ohain." This sentence uses *era* in its traditional definition as a distinct, historical time period. Participants reported that this sentence was eighty-one percent acceptable. They also reported a fifty percent chance of using this sentence.

The second sentence used *era* in its more current definition as a phase or stage in one's personal life. The sentence was also found from the NOW corpus and was "Maybe it's just me, but my lonely era was also my 'treat yourself' era." Participants ranked this sentence as seventy percent acceptable. Forty-three percent of surveyees reported that they could see themselves using this sentence.

Discussion

I hypothesized before conducting the study that *era* had already begun to shift to a slang term, but was not yet widespread, nor near replacing the original definition of the word. All results, both the corpus search and survey, provide evidence of the truth of this presumption.

The NOW corpus search showed that [my * era] remained a steady, minimal part of colloquial language until Swift announced her tour in 2022, when the words per million spiked. Then the following year, it doubled as the tour started and Swift completed the first US leg (Ray, 2023). As Swift continues her tour globally in 2024, it would not be unexpected to see [my * era] usage continue to increase. For future research, it would be interesting to search [era] without "my *" to see if awareness and general usage has increased as well since people are hearing the word more due to the tour.

In the study, the average ranking of *era* usage in conversation before the tour announcement was 14.19. After it was announced, people ranked the average usage as 23.66, resulting in a 9.74 difference. This is not an incredible or significant increase, but an increase in usage nonetheless. *Era* is not the most common word in casual conversation, therefore this slight increase is in line with what could be expected. This does not necessarily demonstrate a shift in the definition of *era*, just an uptick in usage which correlates with *era* becoming a colloquial slang term. The rest of the article will cover the change in definition

Era was not considered a slang word by sixty percent of participants. Though the survey did not ask what they would have reported before the Eras tour, arguably, the number would presumably have been significantly increased considering the word's general increased usage in conversation. This would be another interesting aspect to consider for future research. It is difficult to claim that *era* is now a slang term with only forty percent of participants supporting this claim, though it is clear that the word is on its way there. This state of change was accurately surmised in the hypothesis.

Participants ranked the sentence using era with its new definition as seventy percent acceptable. This is only ten percent less acceptable than the aviation sentence with *era*'s traditional usage. This proves that in addition to *era* being used more frequently,

it is also being used in the context of its new meaning more frequently as well, proving the hypothesis to be correct. *Era* in its new slang form is becoming more widespread and acceptable to the general public.

A final piece of future research to consider would be the pronunciation of *era*. Both common pronunciations of *era*, [ira] or [era], are Standard American English (Oxford 2023). Swift pronounces the word as [era]. An updated survey or an extended survey could assess how people have pronounced *era* in the past and currently, to see if there is a change to match Swift's pronunciation.

Conclusion

Taylor Swift has certainly changed many aspects of language, including expanding the definition of a word with an entire history and influencing its rapid spread. As *era*'s usage is increasing, it is on the brink of being considered a slang term. The definition has also shifted from a traditional distinct period in history to a stage of a personal life. Understanding this single shift caused by one musician can help us understand how pop culture icons influence language and the weight their personal usage has on the rest of the world.

References

Cavender, Elena. (2022). It's not a phase. It's an era. *Mashable*. https://mashable.com/article/era-trend-tiktok

Kopstein, Jeannie, & Mariah Espada. (2023). The staggering economic impact of Taylor Swift's Eras Tour. *Time*. https://time.com/6307420/taylor-swift-eras-tour-money-economy/

Luu, Chi. (2020). The linguistic evolution of Taylor Swift. *JSTOR Daily*. https://daily.jstor.org the-linguistic-evolution-of-taylor-swift/

Oxford University Press. (n.d.). Era. *Oxford advanced learner's dictionary*. https://www.oxfordlearnersdictionaries.com/us/definition/english/era

Ray, Michael. (2023). Taylor Swift. *Britannica*. https://www.britannica.com/biography/Taylor-Swift

Siagallan, Sari Rishita, Sulastri Manurung, & Juwita Boneka Sinaga. (2017). Analysis of figurative language and imagery in Taylor Swift's songs. *Journal Anglo-Saxon*. *8*(1). https://doi.org/10.33373/anglo.v8i1.984

Swift, Taylor. (2020). Illicit affairs [Song]. On *Folklore*. Republic.

Appendix

Survey Given to Participants

Hello! Thanks for clicking on the link! My name is Alayna and I am studying the usage change of the word *era* in the context of Taylor Swift for a class. This survey is purely academic and your responses will be kept confidential. This survey should not take more than five minutes.

Question 1: How old are you?

10–15

16–20

21–30

31–40

41–50

50+

Question 2: What is your gender?

Male

Female

Non-binary/third gender

Prefer not to say

Question 3: Do you consider yourself to be a Swiftie?

No

Yes

Note: Taylor Swift announced her Eras Tour on Nov. 1 2022.

Question 4: How often did/do you use *era* in casual conversation?

Before Taylor Swift announced her tour: Scale from 0–100

After Taylor Swift announced her tour: Scale from 0–100

Question 5: Do you consider *era* to be a slang word?

No

Yes

Question 6: Sentence: A whole second era of aviation was opened up by Whittle and yon Ohain (Smithsonian 2003).

How acceptable is this statement: Scale from 0–100

Could you see yourself saying or writing this sentence: Scale from 0–100

Question 7: Sentence: Maybe it's just me, but my lonely era was also my "treat yourself" era (LA Times 2023).

How acceptable is this statement: Scale from 0–100

Could you see yourself saying or writing this sentence: Scale from 0–100

Collocation Analysis of Female-Related Words in General Conference

Isabel Tueller

This article analyzes how attitudes towards women in the Church of Jesus Christ of Latter-day Saints have changed over time through a collocation analysis of the adjectives used to describe women and other female-specific words in general conference. The article finds that female-specific words collocate most strongly to age-related adjectives and familial status adjectives, though their strength of association has decreased over time. This decrease, combined with the introduction of new collocations in the more recent decades, shows a potential change in attitudes towards women in the Church.

A collocation analysis can tell us much about the attitudes toward a specific noun or the group it refers to. In this paper, I will use a collocation analysis to determine and discuss attitudes toward women in the Church of Jesus Christ of Latter-day Saints, using a corpus of the Church's general conference talks. General conference is a collection of talks by leaders of the Church meant for the whole Church, so its use of language should be representative of Church attitudes as a whole. With the introduction of more women's leadership positions within the Church in the past ten years and the expansion of women's roles in the workforce outside the Church over the past seventy-five years, I am specifically interested in how the adjectives used to describe women have changed over time. Other methods, like surveys of Church members or simple topic analyses, could be used to evaluate attitudes toward women in the Church, but a corpus analysis allows us to use empirical numbers to evaluate attitudes without relying on subject honesty and self-introspection. I will look specifically at female-specific words and their adjectival collocations (or the adjectives found surrounding them) and how their strength of association changed over time in the past one hundred years of talks in general conferences of the Church of Jesus Christ of Latter-day Saints.

Literature Review

The study of collocates, or words often used around each other, of *women* and similar words compared to their male equivalents is not new. In 2008, using the British National Corpus, Pearce found that verbs of physical strength and endurance collocate with *MAN* more strongly than *WOMAN*, as well as adjectives about physical mass (pp. 7–8). *MAN* also more strongly collocates with words surrounding ownership and legal systems (p. 10). *WOMAN* tends to collocate stronger with medical procedures and crimes like rape and abduction (pp. 10–11). *WOMAN* also more strongly collocates with adjectives about class and nationality than *MAN* does (p. 12). The study found many other patterns that followed the stereotypes of men and women (p. 19). Jevric (2017) adds to Pearce's discussion by pointing to how *WOMAN* more often collocates with words related to sexuality and age. In a study of a corpus of books and newspapers, Gesuato (2003) found that "in the discourse practices of English speakers, the world of females is frequently associated with notions of passivity, negativity, and

physicality, while that of males with notions of activity, positivity, and cognitivity." All of these studies show that how society views its stereotypical man and woman correlates with the common collocations of *MAN* and *WOMAN*.

There have also been several studies and research papers done about changing attitudes toward women in the Church of Jesus Christ of Latter-day Saints over the past fifty years, though few studies have been done through a collocation analysis. In 1990, Iannaconne and Miles studied the correlation between changes in women's roles outside the Church (marriage rate, education, occupations outside the home, etc.) and how much the Church (specifically through *Ensign* articles) was talking about women and in what way. A similar study more recently coded articles in the *Ensign* and *Improvement Era* (a Church magazine that ran from 1897 to 1970) as "traditional," "expanded traditional," or "non-traditional" and found a steady increase of nontraditional articles since 1910, with a dip in the 1960s and 1970s (Vance, 2002). In 2008, Hollist compared Ensign articles about women to Mormon women's perceptions of themselves and the ideal Mormon woman.

Studies haven't just looked at the Church magazines. These studies surrounding women within the Church have also been specific to general conference. Tenney (2017) looked at how many times women and female-specific words are referenced in general conference, finding that references dramatically increased right around when the Equal Rights Amendment was passed in Congress (in 1972) and when the General Women's Session was introduced (in 1978). Another study found an increase, though slight, in the number of times women have been quoted in general conference over the past ten years compared to the last fifty years (Wells, 2020). In 2022, Wells built on her previous study, looking at the contexts of when women are quoted in general conference. Part of the study looked at the adjectives describing these quoted women, and Wells found that the quoted women are far more likely to be referred to as *precious, dear,* or *beautiful* than quoted men are. I will build on this small study, expanding it from adjectives used to describe these quoted women to all adjectives near any female-specific words and researching how these adjectives have changed over the past one hundred years.

Methodology

I used the LDS General Conference Corpus (Davies, 2000) to answer my research question. This corpus collects all the general conference talks and sorts collocates by decade. The population of general conference talks does not necessarily perfectly represent the opinions, thoughts, and words of leaders and members of the Church of Jesus of Latter-day Saints. While the talks, speakers, and language of general conference hope to represent the teachings and doctrine of the Church, it still is the more formal and globally focused teachings of the Church, and not perfectly representative. My claims, therefore, can only be applied to the language of general conference and not to the language of all teachings from Church leaders.

Taking inspiration from Tenney's (2017) list of words that he used to find female references in general conference, I will use the lemmas (the dictionary form of a lexeme under which all forms of a word are included) of the following female-specific nouns as search terms: *woman, girl, sister, daughter,* and *mother.* I looked for adjective collocates within four words to include modifiers before the noun and adjectives used over a copular verb to include all adjectives describing the female-specific nouns, whatever role they play in the grammar. I limited my search to the past century, up to the April 2023 general conference talks. I also only included words with a mutual information (MI) score higher than three. The mutual information score is a measure of the strength of association that takes into account how often words occur together as well as how often they occur separately. The majority of words with a collocation frequency of four or less occurred only in one talk, so I collected only words with a collocation frequency of five or higher. Table 1 shows these words.

Table 1

Collocations of female-specific nouns

Rank	Word	Frequency	MI
1	YOUNG	4330	3.21
2	DEAR	1110	4.26
3	BELOVED	696	3.25
4	LOVELY	166	3.21
5	WIDOWED	70	4.65

Rank	Word	Frequency	MI
6	TEENAGE	60	3.69
7	WEEPEST	25	5.36
8	COVENANT-KEEPING	24	4.77
9	THREE-YEAR-OLD	20	3.91
10	UNMARRIED	18	3.55
11	12-YEAR-OLD	15	3.57
12	SAINTED	13	5.00
13	13-YEAR-OLD	13	4.36
14	CORRESPONDING	13	3.26
15	UNWED	12	4.53
16	SEVEN-YEAR-OLD	12	3.56
17	LISPING	11	5.34
18	EIGHT-YEAR-OLD	11	3.02
19	WEDDED	10	4.15
20	CHARMING	10	3.79
21	16-YEAR-OLD	10	3.75
22	HEARTBROKEN	10	3.09
23	10-YEAR-OLD	8	4.38
24	GRIEF-STRICKEN	7	4.03
25	11-YEAR-OLD	7	3.75
26	17-YEAR-OLD	7	3.47
27	COARSE	7	3.42
28	NINE-YEAR-OLD	7	3.32
29	SAINTLY	7	3.15
30	PREGNANT	7	3.07
31	TEN-YEAR-OLD	6	4.05
32	THIRTEEN-YEAR-OLD	6	3.59
33	TEN-	5	4.62
34	EXPECTANT	5	3.62
35	SEVENTEEN-YEAR-OLD	5	3.40
36	TWO-YEAR-OLD	5	3.27
37	SIXTEEN-YEAR-OLD	5	3.04

I felt like this list excluded some of the important collocations I had seen in previous, less specific queries, so I also included collocations with the same parameters of just the lemma of *woman* (see table 2).

Table 2

Collocations of WOMAN

Rank	Word	Frequency	MI
1	YOUNG	3514	4.64
2	MUTUAL	121	4.45
3	LOVELY	53	3.29
4	NATIONAL	52	3.01
5	VIRTUOUS	31	3.03
6	SAMARITAN	29	4.31
7	WEEPEST	25	7.08
8	ELDERLY	19	3.28
9	COVENANT-KEEPING	17	6.00
10	GOD-FEARING	11	3.85
11	WEDDED	10	5.87
12	UNMARRIED	9	4.28
13	STALWART	7	3.14
14	COARSE	6	4.92
15	SILLY	6	4.32
16	MISGUIDED	6	3.83
17	CHARMING	5	4.52

With a combination of these two lists, I had a list of the forty-six words most strongly associated with female-specific words in the past one hundred years. I also took out the collocate *weepest* because in all twenty-five instances of the collocation, it acts as a verb in the sentence, "Woman, why weepest thou?" I then sorted the words by their meaning. Using a mixture of the words' dictionary definitions and the context in which they appear (their concordance lines), I grouped the words into five groups: words related to familial status, age-related words, positive descriptors, negative descriptors, and miscellaneous. I started by taking out the age-related words, which were the most obvious group from

definition alone, and then I sorted out the words related to familial status. From here, I used concordance lines to determine if the remaining adjectives were used positively, negatively, or neutrally. A few of the concordance lines used to determine these groups are discussed in the results section.

After creating the five topic groups, I found the mutual information score for each of the forty-six words in each decade to see how the strength of their association with these female-specific words had changed over time. I used my original query ("woman|-women|girl|girls|sister|sisters|mother|mothers|daughter|daughters") to find adjective collocates within four words, but used no limitation on mutual information or frequency and limited the results to the 1920s. I looked for each of the forty-six words in the results from this search and recorded their mutual information score. Then I repeated this process for each decade.

Results and Discussion

Overall, an analysis of female-specific collocation shows the existence of female-specific adjectives and descriptors in general conference in the past one hundred years. The word forms of *woman, girl, sister, daughter,* and *mother* occur only 42,408 times in the 103-year span from 1920 to April 2023. They make up 0.17 percent of the twenty-five million words in the corpus, which makes the collocation analysis difficult with only a small amount of data. Even with including adjectives within four words and multiple female-specific words, there are only thirty-seven collocates of the female-specific words with a mutual information score above three and a collocation frequency above five. On the other hand, when the female-specific words are replaced with the corresponding male-specific words ("man|men|boy|boys|brother|brothers|father|fathers|son|sons") with the same search parameters, there are only sixteen collocates that fit the qualifications. This difference could just reflect the fact that female-specific words are used less often and less generally than male-specific words in general conference addresses (Tenney, 2017), so there are fewer male-specific words, but it could also point to the fact that we have more adjectives reserved for women and women's roles.

Familial Status Words

The first group of collocates that I observed consisted of words dealing with women's roles as mothers and wives. I grouped these

words into what I called "familial status words." They included the words *widowed*, *unmarried*, *unwed*, *wedded*, *pregnant*, and *expectant*. Jevric (2017) and Pearce (2008) also found these words collocate strongly with female-specific words in their collocation analyses in more general society. Within general conference, there seems to be a general downtrend in the association strength between familial status words and female-specific nouns (see figure 1). The word *widowed* shows this trend most clearly, with an instance of collocation in every decade and an MI that decreases gradually from 5.62 in 1920 to 3.47 in 2020, except for a small spike in the 2000s. One reason for this could be with quoting women and telling stories about women trending upwards (Wells, 2022); talking about women without needing the context of their familial status also seems to be trending upwards, but this finding would benefit from closer study.

The word *wedded* also shows an interesting trend. From the 1920s to the 1990s, it seems that an unmarried woman is more marked than a married woman because only *unmarried* and *unwedded* appear as strongly associated collocates. In the 1990s, we see the introduction of *wedded* as a collocate. The concordance lines show that it almost exclusively appears in the phrase "lawfully wedded as husband and wife." Though it didn't appear as a collocate until the 1990s, the strength of association from the 1990s onward was enough to make it appear as a strong collocate in the forty-six from the whole one hundred years.

Age-Related Words

The second group of collocates that I observed was age-related words. These words included *young*, *teenage*, and *elderly*, as well as all the exact age descriptors (e.g., *three-year-old*, *12-year-old*, etc.). As both Pearce (2008) and Jevric (2017) discuss, it is not rare for female-related words to collocate with discussion of age. This pattern across general society also seems to appear within the Church. I am not sure why only some specific-age descriptors had mutual information scores above three or why some used numerals and some used spelled-out numbers, but across all the age-related words that made it into the top forty-six collocations, there is a general downward trend in the strength of their association. Because the specific-age descriptors drowned out the more general age descriptors, I divided the graph for general age words (*elderly*, *young*, and *teenage*) from the specific-age descriptors (see figures 2 and 3), but both show a general downward trend.
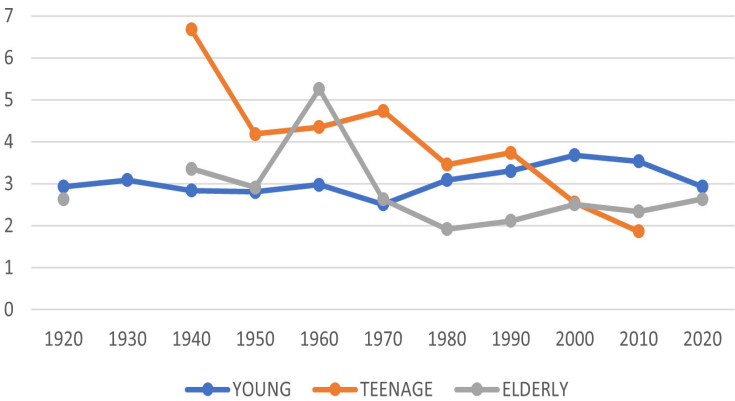
Figure 1

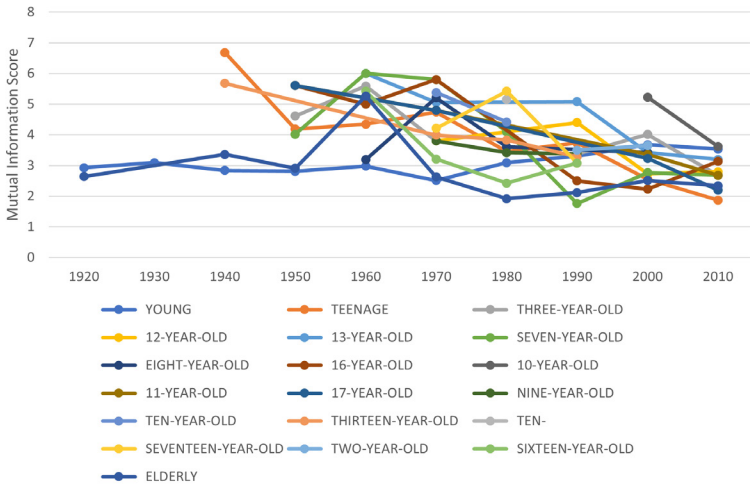*Graphed results of familial status related collocates' MI over time*
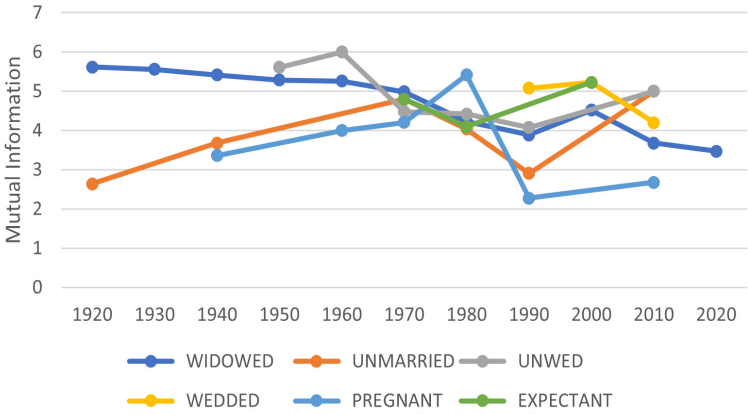
Figure 2

Graphed results of general age descriptors collocates' MI over time

Figure 3

Graphed results of specific-age descriptors collocates' MI over time

Word Sorted by Attitude

From the remaining collocates, I grouped the words into positive, negative, or neither. I sorted them into these groups by looking at how the words were used across their concordance lines. The general trends are not as obvious as the age descriptors and familial status words, but more specific patterns are discussed in this section.

The positive words were *dear, beloved, lovely, covenant-keeping, sainted, charming, saintly, virtuous, god-fearing,* and *stalwart* (seen in table 3). *Dear* and *beloved* were most often used when directly addressing the audience. Many talks begin with phrases similar to "My dear sisters, dear friends, how blessed we are to assemble again," "My dear brothers and sisters, it is a joy to be with you," or "My beloved brothers and sisters, I extend my greetings to all of you." They might end with phrases such as, "Now my beloved brethren and sisters, in conclusion, I bear witness" or offer directives such as, "So to you, dear sisters, I say: Come to Relief Society!" Because collocates within four words were included, you will often see the adjectives appearing before a paired male and female word, like in "brothers and sisters." While the adjective applies to both men and women, it is interesting to note that *dear* and *beloved* do not appear on the collocation list of the corresponding

Table 3

Positive descriptors' MI over time

Word	1920	1930	1940	1950	1960	1970
DEAR	4.13	4.19	3.73	4.47	4.85	4.52
BELOVED	2.49		2.55	3.67	3.28	3.74
LOVELY			2.33	3.81	3.69	3.89
COVENANT-KEEPING						
SAINTED	5.45	5.56	5.68	5.12	6.00	
CHARMING	4.45	4.56	4.68	3.8		4.21
SAINTLY			5.68			3.48
VIRTUOUS	1.36	1.43		1.55		2.08
GOD-FEARING		3.35	2.16	3.44		
STALWART	2.64		1.98	1.91		3.1

Word	1980	1990	2000	2010	2020
DEAR	4.22	3.64	3.87	4.08	4.62
BELOVED	3.30	3.01	3.04	3.07	3.18
LOVELY	3.23	2.97	2.91	3.10	
COVENANT-KEEPING			5.23	4.59	2.88
SAINTED	3.83		5.23		
CHARMING	2.83			5.00	
SAINTLY		2.91	4.23	3.42	
VIRTUOUS	2.89	0.76	1.72	2.18	
GOD-FEARING	2.10	2.50	3.23		
STALWART	1.17	1.76	1.53	3.20	3.88

search with male-specific terms. Male-specific terms seem to be used often enough outside greetings and directives that they do not collocate strongly with words like *beloved* and *dear* that appear in these speech acts.

Some of the other collocates with positive connotations are often used when telling stories about women. *Lovely* was often used in third person when telling stories of women (e.g., "I was stopped by a lovely young woman I did not recognize") or when admonishing men to look for certain types of relationships (e.g.,

"As you date the lovely young women of the Church, you have a duty to protect their physical safety.") *Charming* was used similarly—for example, "Suddenly my eyes met a charming, beautiful young woman." Both were fairly obviously used to describe what the speaker saw as positive traits. Nineteen out of twenty times, *sainted* and *saintly* were used to describe a mother—for example, "I saw and listened to my sainted mother lull her babies to sleep" or "born of a noble father and a saintly mother." In the concordance lines, these words were never used to describe a literal sainted person, but instead used to describe impressive qualities about someone.

One collocate that we see introduced over time is *covenant-keeping*. Before 2000, it does not appear as a collocate but has a strong enough association that it appears in the collocations for the whole century. In its concordance lines, *covenant-keeping* was used to describe a group of women who have certain qualities or standards, women who will receive certain blessings, seen in sentences like, "It is converted, covenant-keeping women whose righteous lives will increasingly stand out," or "As covenant-keeping women, we must shine our gospel light all over the world." *Virtuous*, *stalwart*, and *god-fearing* were used similarly (though only as collocates of the lemma of *woman*, not of all five female-specific words). If the words were associated with receiving blessings or other adjectives with positive connotations, such as courageous, pure, or righteous, I included the collocations as positive. Some example sentences that helped me categorize words as positive include: "Relief Society is for virtuous women, for steady women, for organized women," "I can testify that there are no purer and more God-fearing women in the world, " and "These stalwart women have shown the strength and courage that covenant-keeping women always demonstrate."

One thing I did notice is that *dear* and *beloved* (along with *young*) were consistently in the top four most frequent collocates in each decade, without much change in the strength of their association. This may be caused by the dominance of male speakers in general conference; outsiders of a group feel more need to express love for the group they are outside of. Even with changing attitudes toward women, the expression of love has not changed. I did notice a change in the other word that often appeared in the top four collocates. *Little* was in the top four most frequent collocate of the female-specific nouns from the 1920s to the 1960s and was in the top ten through the 2010s, dropping slowly in both

frequency and mutual information score. Its mutual information score was never high enough to merit a spot on my list of forty-six common collocations.

Table 4

Negative Descriptors' MI over time

Word	1920	1930	1940	1950	1960
HEARTBROKEN					5.42
GRIEF-STRICKEN		5.56	5.68		4.42
COARSE			4.68		
SILLY	3.86		2.88	5.61	
MISGUIDED		2.97			2.19

Word	1970	1980	1990	2000	2010
HEARTBROKEN	2.34	3.10		2.23	3.42
GRIEF-STRICKEN	3.80			4.23	
COARSE	3.48			4.23	4.00
SILLY	3.80				4.00
MISGUIDED		3.42	2.08	3.65	

The negative collocates that I found were *heartbroken, grief-stricken, coarse, silly,* and *misguided* (seen in table 4). *Heartbroken* and *grief-stricken* were the only two found in the collocates of the original search term (with all five female-specific words), while *coarse, silly,* and *misguided* were collocates of only the word forms of *woman*. In six of the ten instances of the collocate *heartbroken*, it is referring to a heartbroken mother whose children were going through hard things or faith challenges. The first instance was in the 1960s, where it occurs only once, but still has a very strong mutual information score. Half of the instances occur in the 2010s, where it also has a strong mutual information score. In the most recent instance (2016, Russell M. Nelson), the speaker is referring to himself as heartbroken about two girls' deaths. It seems like *heartbroken* is a newer word compared to other words in general conference. Research in the future could see if this trend continues in the 2020s. *Grief-stricken* shows the opposite trend. Its first occurrence appears in the 1930s and half of its occurrences appear in the 1940s, where it has the strongest

mutual information score. In each of its concordance lines, *grief-stricken* is used similarly to *heartbroken*: referring to a mother whose child is struggling. The collocation appears after 1950 only in talks by Thomas S. Monson. It seems possible that *grief-stricken* was replaced by *heartbroken*, given their similar usage and almost opposite appearances.

The words *coarse, silly,* and *misguided* are collocates of only the word forms of *woman*. All three words only collocated with *woman* six times each. Four of the six instances of *coarse* come from women's voices. It was often used in contrast to *kind*, like in "There are enough women who are coarse; we need women who are kind," showing its negative use. The other two instances from male speakers refer to the need for men to protect women from coarse behavior. *Silly* comes exclusively from quotes of a scripture in 2 Timothy that talks of "silly women laden with sin" (2 Timothy 3:1), pairing *silly* and the negative effects of sin with women. Four of the six instances of collocation of *misguided* and female-specific words are paired with the corresponding male-specific term, though it does not appear on a collocate search for the word forms of *man*. All three of these more negative terms do not take women or their ideas very seriously. Understanding this possible underlying attitude toward women in general conference may help leaders and speakers choose their descriptors more carefully.

The remaining words (*corresponding, lisping, mutual, Samaritan,* and *national,* seen in table 5) were either ambiguous or neutral in their attitudes, and I sorted them into a miscellaneous group. *Corresponding* and *lisping* were collocates from the original query of all five female-specific nouns, where *mutual, Samaritan,* and *national* were collocates of the word forms of *woman*. All five of the words are found in specific phrases, scriptures, or quotes. In all of its collocations, *lisping* refers to a quote by Abraham Lincoln that says, "Let reverence for the laws be breathed by every American mother to the lisping babe that prattles on her lap." In all of its collocations, *corresponding* precedes the word *age*, pairing either young women or girls with the corresponding male age group. *Mutual*, in almost all of its concordance lines, refers to the Young Women's Mutual Improvement Association. *Samaritan* is used to refer to the story of the Samaritan woman that Jesus talks to at the well. *National* often is used as part of the name of an organization, like the National Woman's Relief Society or the National Council of Women of the United States. Some of the words discussed earlier, like *young, dear,* and *beloved,* also fall into this

category of being used overwhelmingly in set phrases. From this category, it seems that the speakers in general conference often use female-specific words as part of set Church jargon and terminology.

Table 5

Miscellaneous collocates' MI over time

Word	1920	1930	1940	1950	1960	1970
CORRESPONDING			3.68	2.61	5.42	
LISPING	5.45	5.56	5.68		6.00	
MUTUAL	1.72	2.20	3.47	3.98	4.38	2.89
NATIONAL	1.95	2.60	2.63		1.49	1.54
SAMARITAN				3.80	4.83	2.71

Word	1980	1990	2000	2010	2020
CORRESPONDING	4.42				
LISPING					
MUTUAL			1.95	0.30	
NATIONAL	0.28				
SAMARITAN	2.56	2.24	2.65	2.04	1.56

I had expected a clearer change over time in the words used to describe women, but neither the positive nor the negative descriptors had a clear pattern overall. A few of the individual words revealed patterns and meanings, like the introduction of *covenant-keeping*, the relationship between *heartbroken* and *grief-stricken*, and the usage of words like *lovely* and *charming*. The miscellaneous words didn't show attitude so much as they showed the set phrases that we have surrounding women and girls. Further study and data may be needed to draw clearer conclusions.

Conclusion and Application

Just as Jevric (2017) and Pearce (2008) used their studies about the collocations of male- and female-specific words to draw conclusions about the stereotypical man and woman, so too can a study of the collocation of female-specific words in a more specific culture tell us about the generalizations and stereotypes about

women within that culture. This study of the collocations of female-specific words in the Church of Jesus Christ of Latter-day Saints' general conference from the past one hundred years has shown that female-specific words collocate most strongly to age-related words and words related to familial status, just as both Jevric's (2017) and Pearce's (2008) studies also show. This tells us that culture within the Church of Jesus Christ of Latter-day Saints seems to have the same stereotypes surrounding the age and family-centered roles of women. However, the strength of association in both of these collocation groups seems to be trending downward in general conference addresses. As Wells (2020) and Tenney (2017) show, references to women in general conference are on the rise. This may contribute to the weakening of association that we see beginning between stereotypical women categories and female-specific nouns. As women are mentioned more as sources of revelation and guidance and not just as female characters to be taught or talked about, their descriptions become more general and less tied to specific domains. This allows people in the Church to focus less on the stereotypes and more on the doctrine.

Many of the collocates of female-specific words from the past one hundred years do not take women very seriously, focusing on characteristics of age, familial status, emotion (e.g., *heartbroken*). With an increase in collocation strength of words like *covenant-keeping* and a decrease in collocation strength of words about age and familial status, we may be beginning to see a shift in attitudes toward women in general conference talks. Other collocates put women and female-specific words within set phrases and jargon.

References

Davies, Mark. (2000–present). The LDS General Conference Corpus. Available online at https://www.lds-general-conference.org/.

Gesuato, Sara. (2003). The company *women* and *men* keep: What collocations can reveal about culture. *Proceedings of the Corpus Linguistics 2003 Conference*, edited by Dawn Archer, et al., Lancaster University Centre for Computer Corpus Research on Language, 253–62.

Hollist, Julie. (2008). *The ideal Mormon woman: An analysis of* Ensign *articles and comparison to LDS women's perceptions of gender role expectations* [Master's thesis, Utah State University]. https://doi.org/10.26076/344b-51c6

Iannaccone, Laurence R. & Carrie A. Miles. (1990). Dealing with social change: The Mormon Church's response to change in women's roles. *Social Forces*, *68*(4), 1231–1250. https://doi-org.byu.idm.oclc.org/10.1093/sf/68.4.1231

Jevric, Tamara. (2017). 'Medical men' and 'mad women'—A study into the frequency of words through collocations. *Journal of Literature, Culture, and Literary Translation*, *1*(8). https://doi.org/10.15291/sic/1.8.lc.2

Pearce, Michael. (2008). Investigating the collocational behaviour of man and woman in the BNC using Sketch Engine. *Corpora*, *3*(1), 1–29. https://doi.org/10.3366/E174950320800004X

Tenney, Devon. (2017, April 26). *Speaking of gender, generally: Analysis of gendered references and speaking opportunities in LDS General Conferences*. [Poster Presentation]. FHSS Mentored Research Conference, Provo, UT, United States. https://scholarsarchive.byu.edu/fhssconference_studentpub/300

Vance, Laura L. (2002). Evolution of ideals for women in Mormon periodicals, 1897–1999. *Sociology of Religion*, *63*(1), 91–113.

Wells, Eliza. (2020). Quotation and gender in General Conference. *Exponent II*. https://exponentii.org/blog/quotation-and-gender-in-general-conference/

Wells, Eliza. (2022). Quoted at the pulpit: Male rhetoric and female authority in fifty years of General Conference. *Dialogue*, *55*(4).

https://www.dialoguejournal.com/articles/quoted-at-the-pul-pit-male-rhetoric-and-female-authority-in-fifty-years-of-gener-al-conference/

Language Policies and Planning in North Korea

Sydney Christley

This article provides a brief overview of language policies in North Korea including how, when, and why they have been implemented, as well as how they have functioned to strengthen the North Korean nation. The purposes of these policies are outlined including encouraging nationalistic pride, instilling political ideologies, and increasing North Korea's ability to interact and compete at the global level. This article also describes how unofficial assumptions about language, based on the influences of history and international relations, have affected North Korean language policies. A scarcity of information and primary sources from North Korea complicates the study of these policies, but it is evident that the policies have been successful in their mission.

This article provides a brief overview of language policies in North Korea,[1] including how, when, and why they have been implemented, as well as how they have functioned to strengthen the North Korean nation. These purposes include encouraging nationalistic pride, instilling political ideologies, and increasing North Korea's ability to interact and compete at the global level. I examine these topics using Schiffman's brief definition of *language policy* as "decision-making about language," along with the following elaboration:

> I think it is important to view language policy as not only the explicit, written, overt, *de jure*, official and "top-down" decision-making about language, but also the implicit, unwritten, covert, *de facto*, grass-roots, and unofficial ideas and assumptions, which can influence the *outcomes* of policy-making just as emphatically and definitively as the more explicit decisions. (Schiffman, 2006, p. 112)

In other words, language does not exist in a vacuum. This article also attempts to describe how official and unofficial assumptions about language, based on the influences of history and international relations, have affected North Korean language policies.

Language Policies and Nationalism

The idea of a single national language of Korea is not new, as the Korean peninsula has been historically monolingual (or at the least mutually intelligible between dialects) for centuries. During the Japanese colonial period (1910–1945) that immediately preceded the founding of the modern North Korean state, Japanese oppressors tried to erase Korean identity and language. Japanese policies included banning the use of the Korean language in schools, censoring Korean-language newspapers, and even forcing Koreans to use Japanese names in 1940 (Lee, 2018). However, this only cemented an even stronger sense of national pride in the Korean language and culture, and the Korean language came to function as a symbol of Korean identity that was closely tied to the independence movement (Lee, 2018).

---

1. In this article, *North Korea* will be used to refer to the nation officially known as the Democratic People's Republic of Korea and *South Korea* will be used to refer to the nation officially known as the Republic of Korea. *Korea* will refer to the Korean peninsula as a whole.

North Korean propaganda claims that Kim Il-sung, during his time as an anti-Japanese resistance leader, not only defied colonial law and delivered speeches in the Korean language but also established the Association for the Restoration of the Fatherland (*Co.kwuk.kwangpok.hoy*) in 1936. One of the products of this organization was the Ten-Point Program, which, among other things, declared that education was to be compulsory, free, and conducted in the Korean language (Kim, 1971; Lee, 2018). Another organization active in promoting linguistic nationalism was the Korean Language Society (*Co.sen.e.hak.hoy*); many of the leaders of this organization were imprisoned for their participation in the independence movement, such as the compilation and publication of the first Korean monolingual dictionary (Lee, 2018). The memory of colonial rule and the sense of nationalist pride would be fresh in the minds of Koreans—and the early leaders of North Korea—in the first few decades after liberation.

During this time (around 1945–1966), North Korea's language policies can be categorized into three general goals as described by Kim (1978). First, the use of *hangul*—the Korean alphabet—instead of Chinese-style characters was promoted by the government. Adult schools and *hangul* schools were established across the nation in 1946, and by 1948, propaganda claimed that illiteracy had been completely eradicated (Kim, 1978). The use of Chinese characters was seen as feudalistic and contrary to socialism because it required a long period of education only accessible by the wealthy; using *hangul* made reading more accessible to a larger population (Kaplan & Baldauf, 2011).

This led to the second general goal of nativization, primarily focused on vocabulary. Sino-Korean words and phrases (derived from Chinese roots), as well as loanwords from Japanese, Russian, and English, were replaced by pure Korean words (Kaplan & Baldauf, 2011). This language ideology also sought to create new scientific and technical words from Korean word roots (Kim, 1978). This movement reveals a highly conscious awareness of the power of language and a prescriptivist desire to use language to unify the North Korean people. The third general goal of prescriptivism codified these modifications to the Korean language, even directing that children and places should be named with native Korean words (Kim, 1978; Kaplan & Baldauf, 2011). This new way of speaking and writing was called *munhwa-eo* (cultured

language) and was declared the official dialect of North Korea. In January 1964, Kim Il-sung specifically assigned linguists to complete these modifications and update dictionaries and textbooks with the new vocabulary of *munhwa-eo* (Kim, 1978).

Language Policies and Political Ideologies

The state ideology of North Korea, *juche*, which was originally developed from the ideas of Marx and Lenin, was officially adopted in 1972. It emphasizes political, philosophical, military, and economic independence. Although *juche* is most commonly translated as "self-reliance," the work of some scholars asserts that it is, in reality, much more nuanced (Lee, 2003). Kim Il-sung used this policy as a way to reconstruct a war-torn and impoverished country through ideological, economic, and military independence. Throughout the regime of the Kim family, it has cultivated a sense of nationalism and pride in traditional Korean culture, such as music and entertainment. The analogy taught in public schools is that the Great Leader is the brain of the nation, the Workers' Party is the nervous system, and the people are the bones and muscles of the body. This divides the role of each group into the ones who make decisions and give orders (the Great Leader), the ones who channel information (the Party), and the ones who complete those orders (the people) (Lee, 2003, p. 111). Such a system of organization neatly pins the success of the nation on the leadership of one authoritative figure and the support and labor of the masses.

Juche therefore is a vital ideology in North Korea, and its support by the common people must be assured, something that Kim Il-sung and his successors were well aware of. They explicitly used language ideologies like *munhwa-eo* and the nativization of the Korean language in order to accomplish this purpose. The following statement by North Korean linguists clearly reveals this intention:

> We should remake society, nature and remold the people by means of language in order to realize and maintain the independence of the masses…. In remolding the people, one of the most important things is educating and raising the people into communism with Juche style…. It is unthinkable that we can educate the people in the Juche style and inculcate Juche thought in all realms of society without using language. Therefore, language is a powerful weapon

that actively serves ideologically educating and remolding the masses of the people into Juche-style humans. In this context, language carries out the function of ideological education…. Without using spoken and written language, we are not able to learn, research, interpret, and propagate the great Juche ideology. (Choe & Pak, 2000, pp. 75–76; originally published 1999)

Juche was indoctrinated into the North Korean people from birth in the form of education (Lee, 2003) and mass media (Lee, 2018). Lee examined how *juche* ideology was disseminated by the North Korean press during the rule of Kim Jong-il by analyzing data from two corpora: first, a corpus of North Korean newspapers dating between 1998 and 2004 that consists of 2,082,451 words and is funded by the South Korean government; and second, a personally-created corpus of 43,110 words that consists of articles, pulled from the North Korean newspaper *Rodong Sinmun,* which use the word *en.e* (language) from January 1 to September 3, 2014. Data from the first corpus showed that the most common concordances (80.5%) of the word *en.e* referred to categories labeled by the author as "single-nation ethnicity and single-language belief," "language purification (language refinement)," and "superiority of Korean language and ethnicity" (Lee, 2018, p. 68). Data from the second corpus showed that the most common concordances (40.4%) of the word *en.e* referred to categories labeled by the author as "single-nation ethnicity and single-language belief" and "language use in socialism, including language purification" (Lee, 2018, p. 76). These phrases demonstrate the encouragement of language nationalization, language purification, and a homogeneous national and ethnic identity, showing how North Korean language policies and the political ideology of *juche* connect and support each other.

Language Policies and Increasing Globalization

Despite this movement to nativize Korean, Kim Il-sung kept in mind the ultimate goal of reunification with South Korea and made several decisions focused on maintaining mutual intelligibility between the two nations (Kim, 1978). He rejected an attempt in 1964 to revise the Korean alphabet and directed that public education should still instruct students in Chinese characters and

the English language (Kim, 1978). This can be seen as an attempt to satisfy the contradictory demands of nationalism and international participation.

Initially, Kim Il-sung's government had emphasized Russian as the target foreign language that citizens should aim to learn. However, as mentioned earlier, by 1964, English instruction was favored in schools (Kim, 1978; Kang, 2020). American capitalism had become the main enemy of North Korea, and English had become a global *lingua franca* (Kang, 2020). If North Korea wanted to be able to engage in diplomatic and trade relationships with Western countries, as well as countries unaligned with socialism and communism, English language abilities would be required. At this point, English had also become the *lingua franca* of science and technology, which were focal points in *juche* ideology of economic and military independence. As the Soviet Union declined, Russian language abilities became less desired. By the 1980s, English occupied approximately seventy percent of foreign language teaching in North Korea, with Russian at around twenty percent or less. By the 1990s, English language teaching occupied about eighty percent and Russian twenty percent or less, and by 2013, English made up nearly one hundred percent of all foreign language classes in North Korea (Kang, 2020).

Kim Jong-il and Kim Jong-un adopted various policies in order to increase English language abilities in the North Korean population. Kim Jong-il's government sought assistance from organizations such as the Foreign and Commonwealth Office of the United Kingdom and UNESCO, who provided financial support to introduce English as a subject in primary schools as well as secondary schools, improve the language competence of English teachers, send English teachers to international educational conferences and workshops, and revise English textbooks and curricula (Bae, 2015).

Kim Jong-un, in his turn, undertook a revolution of the nation's educational system, seeking to (1) promote economic development by advancing science and technology and (2) strengthen socialist values and suppress those of capitalism (Kang, 2020). Russian language education was completely removed from school curricula, and English language education was significantly bolstered. It was extended to rural areas, where it had been previously unavailable, and was increased by approximately 318 hours on average per student during their primary and secondary school years.

Curricula and textbooks were again revised based on the advice of international organizations such as UNICEF and the British Council, as well as the modern theory of communicative language teaching (CLT) (Bae, 2015).

Although this evidence may seem to indicate a growing willingness in the North Korean population to reverse the historical trend of isolationism, the North Korean government continues to make it clear that the ultimate goal of this language policy is to create a strong socialist nation by improving its citizens' education, instilling *juche* ideology, and strengthening loyalty to the Kim family (Education Committee of North Korea, 2013). Because English language abilities inherently allow more contact with foreign nations and individuals, they are considered a necessary evil for the economic goals of *juche*. The desire for a nationalism—and isolationism—based political system is at the same time supported and threatened by English language education: participation in international science and technology strengthens the nation, but also creates an avenue through which outside influence can enter. This is something that Kim Jong-un's regime is well aware of. As a result, it is not only strengthening political indoctrination in public schools, but co-opting English teaching in order to promote its political ideologies (Kang, 2020).

Conclusion

This article has examined language policies in North Korea throughout its seventy-plus years of existence as an independent nation and the various purposes such policies have served. Pride in the Korean language and culture during the Japanese colonial period easily evolved into extreme nationalism during the Kim regime, which used language policies based on the purification and superiority of the Korean language to encourage such ideas. Language policies also functioned to instill political ideologies such as *juche* and loyalty to the Kim regime. Later, as it became clear that North Korea needed citizens with English language abilities in order to advance in science and technology, the government adopted language policies that would create such citizens. However, they were aware that such a policy could contradict their nationalist ideologies. As a result, they made sure to emphasize that English language abilities both contributed to the welfare and independence of the nation and increased political indoctrination (Kang, 2020).

This research has implications in several fields, including Korean studies, linguistics, political science, and anthropology. The effects of these language policies on the national, communal, and individual levels are still being studied and will probably be studied for as long as such policies exist. The scarcity of information and primary sources available from North Korea complicates this process. Yet, it is clear that these policies (along with other government policies in various areas of nationalism, isolationism, militarism, and socialism) have served the purpose of creating a united North Korean identity—and, arguably, succeeded.

References

Bae, J. (2015). *A study of the English language education of North Korea in the 2000s*. Unpublished Doctoral thesis. Seoul: University of North Korean Studies.

Choe, C-H., & Pak, C-S. (2000). *Cwucheycek en.e lilon yenkwu* [Studies on Juche language theory]. Seoul: Pak.i.ceng. (Original work published 1999 in Pyongyang: Sahoy Kwahak Chwulphansa [Social Science Publishing Company]).

Education Committee of North Korea. (2013). *School curricula and teaching guidelines of overall 12-year compulsory education* (high school). Pyongyang: Education Committee of North Korea.

Kang, Hyeon-Seok (2020). Changes in English language policy in Kim Jong-un's North Korea: A prelude to reconciliation?. *English Today*, *36*(1), 30-36. https://doi.org/10.1017/s0266078419000191.

Kaplan, Robert & Richard Baldauf Jr. (2011). North Korea's language revision and some unforeseen consequences. *Handbook of language and ethnic identity*, 153-167.

Kim, Chin-Wu (1978). Linguistics and language policies in North Korea. *Korean Studies*, *2*(1), 159-175. https://doi.org/10.1353/ks.1978.0002.

Kim, Il-Sung (1971). *Revolution and socialist construction in Korea*; Selected writings of Kim Il Sung. New York: International.

Lee, Grace (2003). The political philosophy of Juche. *Stanford Journal of East Asian Affairs*, *3*(1), 105-112.

Lee, Jae Sun (2018). *State ideology and language policy in North Korea: An analysis of North Korea's Public Discourse* (Doctoral dissertation, University of Hawai'i at Manoa). Retrieved from https://byu.idm.oclc.org/login/?url=https://www.proquest.com/dissertations-theses/state-ideology-language-policy-north-korea/docview/2190677606/se-2

Schiffman, Harold & Thomas Ricento (2006). Language policy and linguistic culture. *An introduction to language policy: Theory and method*, 112.

Detecting Semantic Differences between LDS and Christian Speech

Peter Williams

This article compares usage and meaning of religious language between English-speaking members of the Church of Jesus Christ of Latter-day Saints (LDS) and other English-speaking Christians. Using speech collected from Christian and LDS subreddits and LDS speeches, keyness analyses are used to find target words characterizing each group. For each target word, contextual word embeddings (high-dimensional vectors representing the semantics and context of each word) from BERT (Bidirectional Encoder Representations from Transformers) are generated and reduced to two dimensions using Principal Component Analysis (PCA) to visually examine the clusters indicating polysemy correlating with religious groups. This examination finds that some words (e.g., church, sin, Mormon) have visible differences that are detectable using this method. By understanding these differences, speakers can improve mutual understanding between these groups.

Many people believe that Utahns and members of the Church of Jesus Christ of Latter-day Saints speak "weirdly," with a dialect influenced by their region and their religion. While research has been done on some of the differences between Utah English and Standard American English (Bowie, 2003; Eddington and Savage, 2012), there is not much research looking specifically at speech from members of the Church. With this in mind, this article endeavors to answer the following questions: Which words are used more frequently in Latter-day Saint speech, and how do keywords in Latter-day Saint speech compare to those of other Christian denominations? Additionally, this article seeks to answer which of the common keywords between Latter-day Saint speech and Christian speech have dual senses: one sense used by members of the Church and another used by other Christians.

Recent research has attempted to automatically detect semantic shifts using contextual word embeddings (representations of both context and semantics). This allows for research into polysemy (multiple meanings for the same word) and shifts between senses of a single type. This is in place of static word embeddings from a model like word2vec, which makes an implicit assumption that all occurrences have the same sense and that there is no polysemy. With the advent of contextual word embeddings (Peters et al., 2018; Devlin et al., 2018), this field shifted to looking more closely at individual words and detecting semantic shifts (Kutuzov et al., 2018). Kutuzov et al. recently evaluated these approaches to improve understanding of their output (2022). They determined that these contextual embeddings aren't always only revealing semantic shifts but also frequently reveal syntactic changes and are sensitive to the frequent use of a word in a very specific context in a particular class. For this article, I am ignoring the potential conflation of differences between semantic and syntactic changes because the goal is simply to determine if there are any differences—regardless of type—between religious subgroups.

There is some precedent for studying religious language, although it primarily looks at religious language regardless of denomination compared to more general speech (Ellis, 2021). I have found little research comparing language usage between specific religious groups, especially subgroups of a certain religion, such as various denominations of Christianity. This comparison between members of the Church of Jesus Christ of Latter-day

Saints and members of other Christian denominations would fill gaps in research about how religious language differs between religious subgroups.

Methods

Corpus Creation

To answer the proposed research questions, I first collected a corpus of speech specifically from members of the Church of Jesus Christ of Latter-day Saints from various sources. This includes a corpus of more than twelve thousand talks from the general conferences of the Church, combined with BYU devotionals and comments and posts from predominantly Latter-day Saint subreddits on Reddit. In total, the corpus of Latter-day Saint speech is over thirty-four million tokens. Only the three million tokens from Reddit were used for word embeddings and evaluation due to computing limitations and for comparability to the Christian corpus (in total, nearly six million tokens came from Christian subreddits that are not dedicated to any particular denomination).

Corpus Comparison

To select keywords to examine and map with word embeddings, I loaded the corpora into the corpus analysis tool AntConc (Anthony, 2014). I performed several keyness analyses: Latter-day Saint speech against general speech, Christian speech against general speech, Latter-day Saint speech against Christian speech, and Christian speech against Latter-day Saint speech. The metrics I used measured the likelihood that the word occurred in the target corpus instead of the reference corpus and the ratio of relative frequencies between the two. I sorted by effect size as the primary measure of which words were different between the groups. Using this, I decided on a shortlist of fifty words to use to compare word embeddings.

To create the contextual word embeddings, I used Bidirectional Encoder Representations from Transformers or BERT (Devlin et al., 2018). I split every full text into sentences using spaCy's Sentencizer component (Honnibal et al., 2020). Words were then tokenized by the BERT Tokenizer and padded or truncated to a max length of sixty-four. The tokens were passed through the model to create tensors of 12 hidden layers and 768 features. After

generating contextualized word embeddings for both corpora, I used scikit-learn's implementation of Principal Component Analysis (PCA) to project the embeddings onto a two-dimensional space and then plotted the distributions using Matplotlib (Pedregosa et al., 2011) to look for clusters.

Results

Keywords

In evaluating the keywords for each keyness analysis, some interesting trends appeared that helped inform the choices for the word list, and helped explain more of the differences in speech between groups.

Latter-day Saint Speech

In Latter-day Saint speech compared to general speech, four main groups of keywords appeared: place names, people names, doctrinal terms, and organizational terms. Of these, we would not expect names to have significant semantic or syntactic differences from their appearance in Christian speech; we simply do not expect them to appear frequently because they are not relevant to that group. Some examples of names include *Nephi*, *BYU*, *Nauvoo*, *Heber*, and *Nephites*. Although not semantically or syntactically interesting, the names do demonstrate that the history and culture of the separate groups heavily influence what is different about their speech. This makes sense especially when we consider the emphasis members of the Church place on quoting from scripture and general authorities when discussing doctrinal questions.

For the contextual embeddings, I primarily selected from the organizational and doctrinal terms, which are likely to also appear in Christian speech, but may have different connotations, appear in different syntactic constructions, or have some other difference in semantic content. Some examples include *covenant, ordinance, tithing, redeemer, sacrament,* and *gospel* as doctrinal terms in addition to *apostle, prophet, bishop, church*, and *temple* as organizational terms. The organizational terms were less likely than the doctrinal terms to appear on the Christian speech keyword list, indicating that it is probable they are not as frequent. Therefore, we may not recognize significant semantic differences because the word is primarily used by one group alone.

It makes sense then that many of the organizational terms are less likely to appear in Christian speech because they are specific to the Church of Jesus Christ of Latter-day Saints, and so other Christians who are not familiar with the specifics of the organization of the Church would not use them. The terms that are shared (*apostle, church*) are more universal for Christian churches, or related especially to the Bible or Biblical church. *Apostle* is one word in particular where I noticed a difference in meaning. The use in Latter-day Saint speech refers to modern prophets, and the use in Christian speech refers to apostles from the New Testament. This is one example of a semantic shift that we may expect to see in contextual embedding, although the uses of the word are similar, despite the time differences.

Christian speech

In Christian speech compared to general speech, four main groups appeared as well: terms from Reddit, references to Bible editions or books of scripture, positive doctrinal terms, and negative doctrinal terms. The first two were ignored because speech specific to the Reddit environment (e.g., *subreddit, mods*) is not our focus, and references to Bible editions/books (e.g., *NIV, NKJV, Corinthians*) are not relevant for semantic differences, although their increased frequencies may indicate another facet in which Latter-day Saint speech distinguishes itself from that of other Christians. The most likely parallel to the Bible editions and books in Latter-day Saint speech is that some of the names that appear (e.g., *Mormon, Moroni, Nephi*) are names of a book of scripture from the Book of Mormon, in addition to being used as a person's name. Positive and negative doctrinal terms include *Jesus, bless, righteousness,* and *spiritually* along with *Satan, sinning, adultery,* and *demonic,* respectively. Both groups were pulled from to be used as keywords to filter for analysis with PCA. The negative keywords were notable because they almost did not appear at all in the Latter-day Saint speech.

Latter-day Saint vs. Christian speech

In the keyness analyses comparing Latter-day Saint and Christian speech to each other, the same patterns noted above continued, with many names and some organizational terms on the Latter-day Saint vs. Christian list, and more negative doctrinal terms, as well as swearing and more Reddit-specific terms, appearing on the Christian vs. Latter-day Saint list. The top fifteen words that

characterize Latter-day Saint speech in comparison to general Christian speech are shown in table 1. In this list, we note names of people (*Hinckley, McKay, Mosiah*), names of places (*Nauvoo*), and specific Latter-day Saint organizational terms (*wards, quorum*). Two additional words specific to Latter-day Saint culture are *ensign* and *deseret*, which have unique history with the Church and are also relevant as the name of a previously Church-published magazine and as part of the name of several brands under the organization of the Church, respectively.

Table 1

Top 15 Latter-day Saint vs. Christian keywords

Keyword	Likelyhood	Effect
ensign	2134.485	2316.996
hinckley	989.065	1049.053
heber	966.371	1049.053
quorum	1817.157	994.682
mckay	1783.476	976.383
mosiah	761.809	826.996
quorums	737.82	800.954
woodruff	690.813	749.927
wards	1352.182	723.885
deseret	666.824	723.885
nauvoo	654.505	710.513
lehi	634.73	689.046
monson	563.086	611.273
bensen	554.333	601.771
hyrum	518.349	562.709

Principal Component Analysis

From examining the PCA scatterplots for the selected keywords, I noticed some patterns. First, many of these graphs do not clearly display multiple clusters that indicate some systematic difference in usage, either syntactic or semantic. Some may have clusters or extensions of clusters that possibly suggest a different usage by one group or the other, but it is not clear that they actually constitute a significant difference. Additionally, some plots simply do not have enough points from one group or the other, or both, to determine if there is a significant difference. However, if it appears that there are multiple clusters of meaning, there is frequently one cluster where LDS and Christian speech overlap. Then there is another smaller cluster of Latter-day Saint speech, suggesting that there is a shared meaning a lot of the time, but in particular contexts, LDS speakers use a word differently.

One example of this is the word *church*, which shows a large spread of points with both Christian and Latter-day Saint speech, and then a separate smaller cluster made almost entirely of Latter-day Saint data points with one Christian point, as seen in figure 1. This second dense cluster indicates that there is an additional meaning (or syntactic structure) that Church members frequently use that is distinct from general Christians. By looking at concordances, I noticed some possibilities for what this second cluster could represent. This could represent the consistent use of the full name of

Figure 1

*PCA projection of contextual word embeddings for the type* church
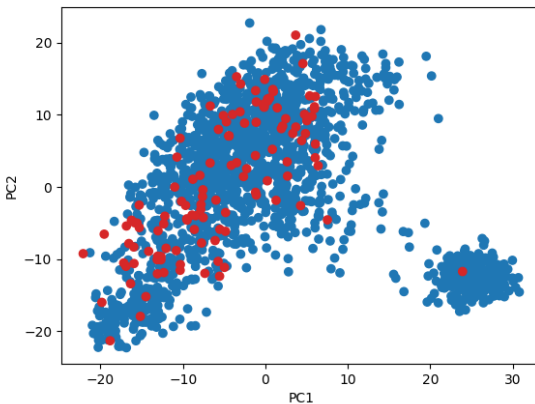
the Church (a possibility supported by similar clusters for *Jesus, Christ,* and *Saints*) or frequent use of *the church* instead of *a church* or *the Catholic church*. This suggests a distinct semantic difference exists when Latter-Day Saints use the word *church* to specifically mean the Church of Jesus Christ of Latter-day Saints, in contrast with the meaning or referent implied by speakers from other Christian denominations. A similar example where the token as used in Latter-day Saint speech likely has a more specific referent is *Joseph*, which would refer to Joseph Smith.

Another example demonstrating a likely semantic difference is the word *mormon* (seen in figure 2). In the tokenized and embedded data, most instances found of the type *mormon* were from Latter-day Saint speech. As such, this is mostly informative about detecting the polysemy of a word even within a single group. The different meanings demonstrated here likely are associated with the Book of Mormon, Mormon as a person, or Mormon as an adjective describing someone who is a member of the Church of Jesus Christ of Latter-day Saints.

Figure 2

*PCA projection of contextual word embeddings for the type* mormon
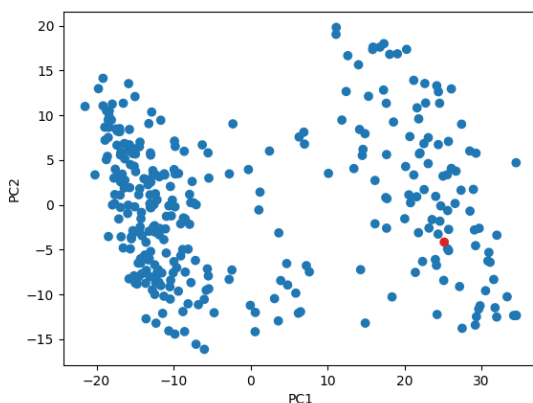
Finally, the plot for the type *sin* is one example contrary to the rest where the Christian speech shows a separate cluster from the main cluster that overlaps between Christian and Latter-day Saint speech, as seen in figure 3. By looking at concordances, it seems likely that this cluster contains instances where *sin* is used

in the phrase *original sin,* which carries a different meaning than when *sin* is used alone. This demonstrates that these contextual word embeddings are likely capable of capturing some level of semantic prosody within a word and not just semantic content of the word itself based on the context.

Figure 3

PCA projection of contextual word embeddings for the type sin
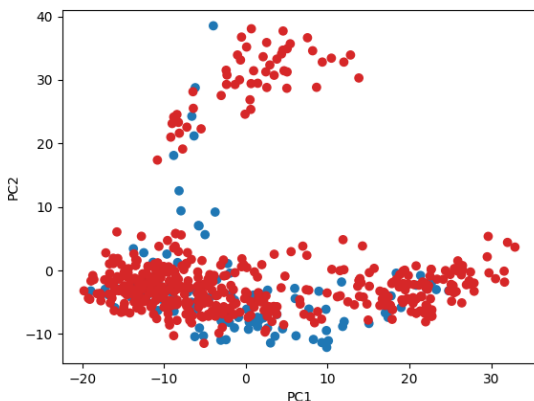
Discussion and Conclusions

Discussion

My results above display several categories of words (and specific examples of keywords within each category) that demonstrate some of the differences between Latter-day Saint and Christian speech. These differences are likely because of organizational, doctrinal, historical, and cultural differences between the Church of Jesus Christ of Latter-day Saints and other Christian churches. Several keyword analyses were able to help determine which words were the most significant in separating the two groups and which categories of words bind the two groups together under the broad generalization of "religious language." This answers the first research question about determining which categories of words distinguish Latter-day Saint speech from Christian speech.

The second research question was answered by using contextual word embeddings for certain keywords and Principal Component Analysis. I determined that for some words, there are significant

differences between their semantic content or syntactic usage in Latter-day Saint speech and in Christian speech. With qualitative analysis of the concordance lines of these keywords, I suggested potential different senses for these words in the two different contexts. The success of these methods in determining that some different senses exist for keywords shared by both groups demonstrates that this process can be applied more broadly to religious or social subgroups that may share words with different meanings, in addition to being used in diachronic semantic shift studies.

These two lists can potentially be applied to those learning about the Church of Jesus Christ of Latter-Day Saints or other Christian churches. Members of those churches can use these lists to be aware of which words might be unknown or have a different sense for an outsider. This could also be applied like a word list for an L2 learner (where the L2 learner is learning about a given church) by giving them a glossary of words that might be used frequently by members of that church.

Future Work

There are several opportunities for future iterations of this research to improve. First, they can start by finding or gathering a corpus of Christian speeches comparable to the corpus of speeches from members of the Church of Jesus Christ of Latter-day Saints so that the two sub-corpora are completely comparable. Additional corpus improvements could include generating contextual embeddings for the entire corpus or using a sliding or larger context window to work around the limiting factor of each contextual embedding using just the sentence as context. Future research could also experiment with fine-tuning BERT on the data first.

The analysis of the embeddings could be improved by using different clustering algorithms and some of the statistical methods and measures from prior research to determine if the difference is statistically significant. Additionally, due to poor formatting of the data after generating embeddings, the full context of each token was no longer connected to the token and embedding itself. As such, I was unable to directly examine the contexts of instances that were in separate clusters, and instead used general KWIC analyses in AntConc to determine what could be a potential cause of the secondary cluster. Connecting the context to the

embedding would be more informative for constructing full sense inventories.

This same analysis could also be done in different languages for the same groups to determine if the same words are key for each group in different languages or if there are cross-language differences. I expect that there would be some differences in the distributions of contextual embeddings, but I don't know what differences there are for keywords in other languages.

Conclusions

I successfully determined which words are key in making Latter-day Saint speech different from Christian speech by using keyness analysis to compare the two corpora to each other and to a general corpus. This answers the first research question with a list of some words that most distinguish Latter-day Saint and Christian speech from each other and from Standard American English, including the use of Church-specific organizational and doctrinal terms in Latter-day Saint speech and more negative doctrinal terms in Christian speech. The second research question was answered by proving the viability of the application of diachronic semantic shift detection methods (contextual word embeddings with PCA) to synchronic detection of semantic differences between two social groups of speakers of the same language. For this study, there were certain words (*church, sin, jesus, mormon*) that demonstrated these semantic differences. Overall, I determined that there are usage and semantic-based differences between Latter-day Saint and Christian speech, which can be determined using corpus-based and computational approaches. Speakers should be aware of those differences to increase mutual understanding despite linguistic differences.

References

Anthony, Laurence. (2014). *AntConc (Version 3.4. 3)[Computer Software]. Tokyo, Japan: Waseda University.*

Bowie, David. (2003). Early development of the card-cord merger in Utah. *American Speech, 78*(1), 31–51. https://doi.org/10.1215/00031283-78-1-31

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, & Kristina Toutanova. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv Preprint arXiv:1810.04805*. https://doi.org/10.48550/arXiv.1810.04805

Di Paolo, Marianna, & Alice Faber. (1990). Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language Variation and Change, 2*(2), 155–204. https://doi.org/10.1017/S0954394500000326

Eddington, David, & Matthew Savage. (2012). Where are the Moun [ʔə] ns in Utah? *American Speech, 87*(3), 336–349. https://doi.org/10.1215/00031283-1958345

Ellis, David Paul. (2021). *A Philosophical Investigation of Religious Language: A Study of the Identity, Meaning and Semantics of Religious Utterances*. Bangor University (United Kingdom).

Honnibal, Matthew, Ines Montani, Sofie Van Landeghem, Adriane Boyd, & Others. (2020). *spaCy: Industrial-strength natural language processing in python*. https://doi.org/10.5281/zenodo.1212303

Kutuzov, Andrey, Lilja Øvrelid, Terrence Szymanski, & Erik Velldal. (2018). Diachronic word embeddings and semantic shifts: a survey. *Proceedings of the 27th International Conference on Computational Linguistics*. 1384–1397. Retrieved from https://aclanthology.org/C18-1117

Kutuzov, Andrey, Erik Velldal, & Lilja Øvrelid. (2022). Contextualized embeddings for semantic change detection: Lessons learned. *ArXiv, abs/2209.00154*. Retrieved from https://api.semanticscholar.org/CorpusID:251868792

Laba, Yurii, Volodymyr Mudryi, Dmytro Chaplynskyi, Mariana Romanyshyn, & Oles Dobosevych. (2023). Contextual embeddings for Ukrainian: A large language model approach to word

sense disambiguation. *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*. https://doi.org/11–19. 10.18653/v1/2023.unlp-1.2

Lee, Sarah, & Debra Ziegeler. (2006). Analysing a semantic corpus study across English dialects: Searching for paradigmatic parallels. *Corpus linguistics around the world* (121–139). Brill. https://doi.org/10.1163/9789401202213_011

Liu, Yang, Alan Medlar, & Dorota Glowacka. (2021). Statistically significant detection of semantic shifts using contextual word embeddings. *arXiv Preprint arXiv:2104. 03776*. https://doi.org/10.48550/arXiv.2104.03776

Mohamad, Tamam. (2023). The status of religion/sect-based linguistic variation in Tartus, Syria: Looking at the nuances of Qaf as an example. *Languages, 8*(3), 167. https://doi.org/10.3390/languages8030167

Montariol, Syrielle, Elaine Zosa, Matej Martinc, & Lidia Pivovarova. (2020). Capturing evolution in word usage: Just add more clusters? *Companion Proceedings of the Web Conference 2020*. Retrieved from https://api.semanticscholar.org/CorpusID:210839389

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, & Édouard Duchesnay. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Peters, Matthew E., Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, & Luke Zettlemoyer. (2018, June). Deep contextualized word representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1*(Long Papers), 2227–2237. https://doi.org/10.18653/v1/N18-1202

Stanley, Joseph A., & Kyle Vanderniet. (2018). Consonantal variation in Utah English. *Proceedings of the 4th Annual Linguistics Conference at the University of Georgia*, 50-65.

Attitude and VOT in L2 Learners

Salem Peckinpaugh

This article examines the effect of social factors, such as ethnic group iden-tification and attitude towards a second language (L2) on the voice onset time (VOT) of the voiceless plosives /p/, /t/, and /k/. Previous research done on this topic has yielded contrasting results, and due to this incongru-ity, this experiment was conducted in order to contribute to the discussion. The participants were sorted into groups based on their experience with Spanish, their attitude towards it, and the VOTs of participants speaking both English and Spanish during voice recordings which were analyzed in Praat. Ultimately, this article finds that attitude towards an L2 does not influence the VOT of the speaker.

Introduction

I t has been well-established that voice onset time (VOT) varies across languages. For example, the VOTs of voiceless plosives in Spanish are significantly shorter than those in English. When someone is learning a second language, they may carry over more similar VOTs from their native language, thus affecting aspects of language such as pronunciation and how "native-like" one may sound. Different researchers have previously shown contrasting results for similar research questions regarding the effects of a learner's attitude on the VOT they subsequently produce in their second language (L2). The research question considered in this case is the following: What is the effect of social factors, such as ethnic group identification and attitude towards an L2, on the VOT of voiceless plosives /p/, /t/, and /k/? It is hypothesized that there will be a positive correlation between the above-stated factor of attitude and the resulting VOT times of speakers with English as their L1 and Spanish as their L2.

Literature Review

On average, the voice onset time (VOT) for the voiceless plosives /p/, /t/, and /k/ in Spanish are, respectively, five milliseconds, ten milliseconds, and thirty milliseconds. In English they are roughly sixty, seventy, and eighty milliseconds (Lisker & Abramson, 1964). Perhaps one of the most thoroughly researched topics related to VOT is the production and perception of accents. In 1995, Elliot conducted a study on the effects of formal instruction on pronunciation in adult L2 learners. He found that while neither attitude nor concern for accuracy yielded significant results on pronunciation, he did find that the formal instruction they had set up for those in the experimental group showed significant improvement in their resulting pronunciation (Elliott, 1995). This study was important for the research for this article because it presented the information that attitude and concern for accuracy did not affect pronunciation. However, much more recently, in 2021, a study was published that found that ethnic group identification with the L2 had a positive correlation with accuracy of pronunciation in the L2 (Edwards et al., 2021). Because of the findings in this study, one of the questions on the questionnaire in this experiment was the participants' affiliations with and exposure to ethnic groups that are associated with Spanish.

In 2015, a study aimed at finding the connection, if any, between VOT and the perception of accents, especially as related to the duration of aspiration. This study shows that speakers who are earlier on in the learning process of their L2 or have not been exposed to immersive L2 environments tend to have more L1 transfer of average VOT time from their native language, which contributes to their perceived accent when speaking (Schoonmaker-Gates, 2015; Gorba, 2020). In the same year, a different study looked at Swedish and Spanish and the production of VOT in L2 rather than the perception of accent because of VOT. It recorded the VOT in milliseconds and did present a correlation between age and VOT such that those who are early L2 learners (starting by age eleven) had more native-like VOTs than those who were late L2 learners (Stölten et al., 2015). This is a helpful factor to keep in mind when analyzing the results of this study so that results can be compared by age group since there has been previous evidence of differences.

In 2022, Jomaa performed an experiment examining the relationship between linguistic identity and both the perception and production of phonemes in an L2. In this case, adult L1 speakers of Arabic who spoke English as an L2 yielded results that showed that "participants with a more positive attitude towards English produced more aspirated voiceless stops than those who favored Lebanese Arabic who, in turn, produced longer voiced stops" (Jomaa, 2022). The differing conclusions from Jomaa and Elliott, as stated in the introduction, indicate the need for further research on the subject of VOT and attitude.

Methods

Information was gathered from fourteen participants for this study: seven from Indiana, four from Utah, one from Missouri, one from Pennsylvania, and one from Nevada. They ranged from fifteen years of age to fifty-eight years of age and all stated their ethnicity as Caucasian. It was difficult to recruit more than a few participants for this study since there was a scheduled time frame to present the results, and the process of extracting and analyzing data for this particular experiment was very time-consuming. All of the data was collected from personal connections as well as from connections through the Spanish department here at BYU. Some participants had zero knowledge of Spanish in order to act as a control group, while other participants had a variety of experience

with Spanish. It was important that the participants had differing experiences with Spanish as their L2 and had a range of ages so that a wide variety of participants could be studied while still being relatively few in number. In order to ensure this, a questionnaire was the first step that participants took in this experiment. This survey included questions regarding where they were raised, age, gender, ethnicity, family background with Spanish, the length of time they have spent learning Spanish, and the context of their learning experience (academic or immersive).

Participants were informed before the next portion of the experiment that their voices would be recorded on a software program called Audacity in order to later extract data from their speech. Next, I conducted interviews in English which specifically asked questions about attitudes towards Spanish. By doing so, I was able to gather both the information needed about social factors and the L1 VOT data needed at the same time. Then I instructed them to read a short passage in Spanish in order to gather data regarding their L2 VOT of the same voiceless plosives /p/, /t/, and /k/. Finally, I analyzed the VOT for both English and Spanish using the computer programs ELAN 6.6 and Praat to perform a spectral analysis in which VOT was measured in milliseconds in each of the recordings. All data was entered into a spreadsheet in an organized fashion in order to perform tests. Afterward, another computer program, Jamovi, was used to perform a statistical test in the form of a linear mixed-effects model to ensure that any differences found were not due to chance.

I pulled demographic information, language background, and attitudes towards Spanish from the answers given in the questionnaire and the part of the interview conducted in English. I sorted the participants into categories accordingly and then compared their VOT in both their L1 and L2 by means of a spectral analysis. If the VOT of participants with a positive attitude towards their L2, regarding it as preferred to their L1, are more similar to the averages of VOT times for /p/, /t/, and /k/ in Spanish than in English, then that will confirm my hypothesis. However, if no relationship is found or a negative correlation is found between positive attitudes towards the L2 and the VOT, then that would negate my hypothesis and cause the null hypothesis to be accepted instead.

Results

In an effort to visualize the data, a separate box plot was constructed for each voiceless plosive. They are shown below in figures 1.1, 1.2, and 1.3.
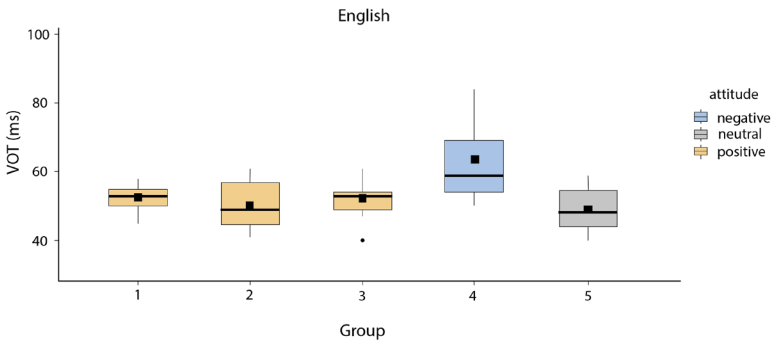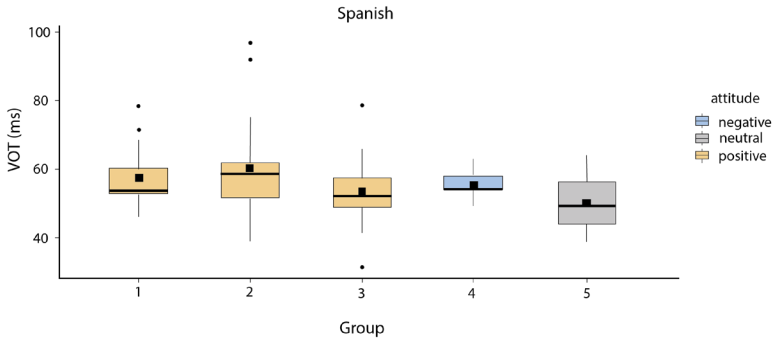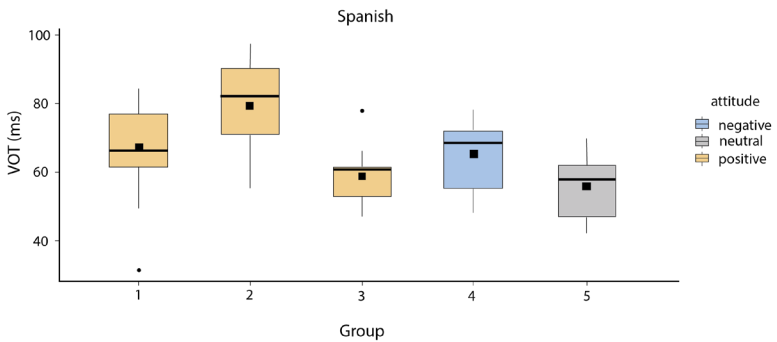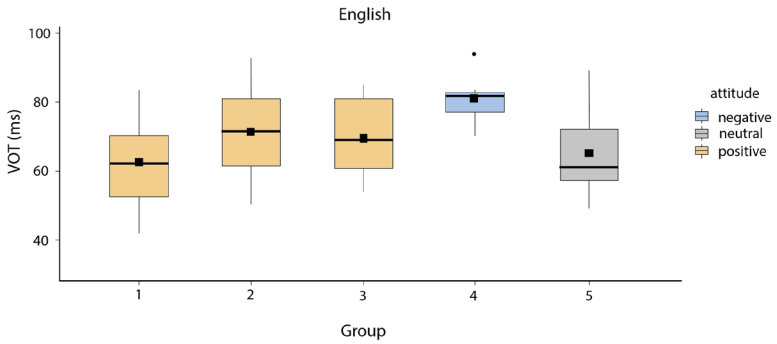
Figure 1.1

(/p/)

Figure 1.2

(/t/)

Spanish

Figure 1.3

(/k/)

English

Spanish
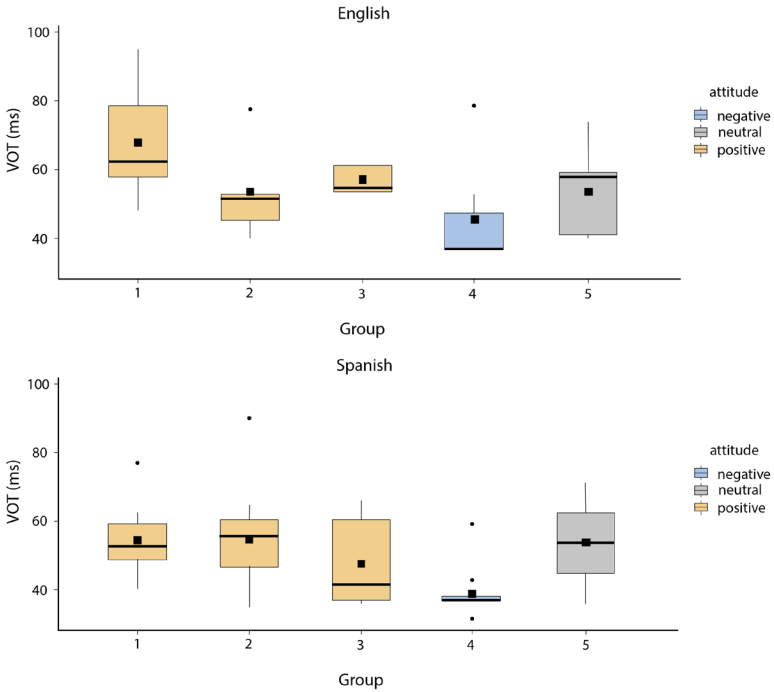
These box plots, chosen due to their ability to visually present both averages and range of results, contribute to addressing the hypothesis by presenting the data in a way that includes each variable present in the study: VOT, language, attitude towards Spanish as an L2, and group. Group 1 consisted of three participants with zero to one year of experience learning Spanish but

with a positive attitude towards the language. Group 2 was made up of four participants with two to three years of experience with Spanish and a positive attitude towards it. Group 3 consisted of two participants with four or more years of experience with Spanish and a positive attitude towards it. Group 4 had two participants, one with two years of experience and the other with six years of experience with Spanish. They were grouped solely by their negative attitude towards Spanish as there were not enough participants with a negative attitude towards their L2 to justify putting them in separate groups based on experience. Group 5 had three participants, one with zero experience, one with two years of experience, and one with four years of experience learning Spanish. They were grouped by the factor of a neutral attitude towards Spanish as an L2 without being split up by experience for the same reason listed above for Group 4. Figure 2 shows VOT as related to group and language and figure 3 shows VOT as related to attitude and language.

Figure 2
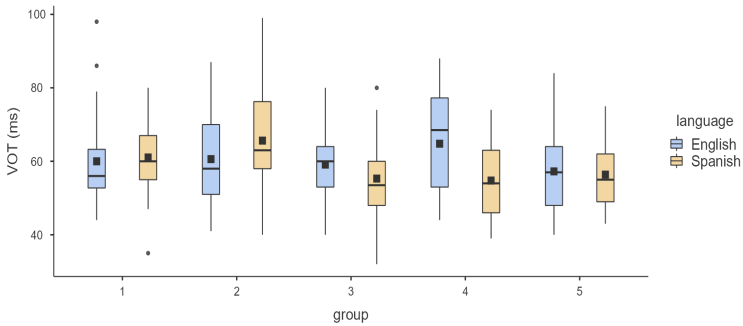
Figure 3
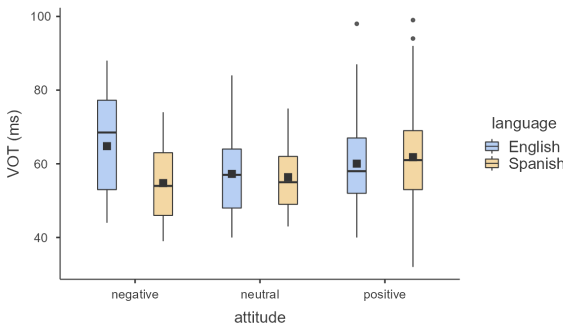
A linear mixed-effects model was performed for the statistical test. This was done in order to determine which factors hold significance. It is important to notice the p-value for each grouping of variable, which shows the level of significance for each. The results are shown below:

Table 1

Fixed Effect Omnibus tests

	F	Num df	Den df	p
attitude	2.03221	2	14.2	0.167
language	0.00120	1	329.1	0.972
phoneme	28.9594	2	325.2	< 0.001
group	1.27312	2	10.1	0.321
language * phoneme	6.78288	2	325.3	0.001
phoneme * group	5.37327	8	324.7	< 0.001

Note. Satterthwaite method for degrees of freedom

Discussion and Conclusion

Based on both the graphs for each phoneme and the statistical test performed above, it can be concluded that there is significance found in the categories of "phoneme," "language and phoneme," and "phoneme and group." The threshold for significance in this study was $p < 0.05$. In the case of the factor of "phoneme," the p-value found was < 0.001. This is evidence that that the VOT is different for each of the phonemes /p/, /t/, and /k/. For "attitude," the p-value found was 0.167. This shows that VOT is not significantly influenced by the factor of attitude. For "language and phoneme," the p-value found was 0.001. This indicates that the VOT for each phoneme is different in English as compared to Spanish. Finally, for "phoneme and group," the p-value was < 0.001. This signifies that the VOT for each phoneme is different depending on the group. As "group" affected the VOT of each phoneme but "attitude" did not, we can conclude by reason of deduction that the influencing factor within "group" was the amount of experience learning Spanish as an L2. Due to these results, the alternative hypothesis is rejected, and instead, the null hypothesis is accepted, which is that attitude towards an

L2 does not affect the VOTs produced. The findings of this study aligned more closely with those of Elliott than those of Jomaa, as discussed in the literature review. The implications of these results are that the more time you spend learning a language, the more closely your VOT of word-initial voiceless plosives aligns with that typical to your L2. Attitude towards your L2, on the other hand, does not appear to make a significant difference. In a real-world application for L2 learners, the findings of this study would suggest that one who hopes to improve their skill level towards sounding like a native speaker should spend more time learning the language, regardless of attitude.

Future Research

If this study were to be recreated for future research, it would be beneficial to gather more participants. Having more participants would allow the researcher to sort them into groups that were separated by a variable such as age, as well as allow them to split up the groups that contained participants with neutral and negative attitudes towards their L2 as in this study; there were too few to justify splitting them up by years of experience with their L2. In addition, if even more time was available to complete the experiment, it could be interesting to look at the percentage of difference of the VOT between English and Spanish for each participant, because even if the overall VOT of one group is not different from another, there may be a greater percent of change for each individual to be found depending on the same factors used in this experiment. Although that was not a test performed in this experiment, it could yield important results that may give reason to reconsider the results of this study which found, in conclusion, that attitude towards an L2 does not influence the VOT of the speaker.

References

Edwards, Jette Hansen, Ka L. R. Chan, Toni Lam, & Qian Wang. (2021). Social factors and the teaching of pronunciation: What the research tells us. *RELC Journal, 52*(1), 35–47. https://doi.org/10.1177/0033688220960897

Elliott, A. Raymond. (1995). Foreign language phonology: field independence, attitude, and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal, 79*(4), 530–542. https://doi.org/10.2307/330005

Gorba, Celia. (2020). Crosslinguistic influence in L2 speakers: The effect of language experience on L1 and L2 VOT production and perception [Doctoral dissertation, Universitat Autònoma de Barcelona]. *Dipòsit Digital de Documents de la UAB*. Tesis Doctorals. https://ddd.uab.cat/record/243180

Jomaa, Danielle. (2022). Linguistic identity and speech: How one's linguistic attitude towards L1 and L2 affects the perception and production of L2 phonemes [Doctoral dissertation, American University of Beirut]. *AUB ScholarWorks*. AUB Students' Theses, Dissertations, and Projects. http://hdl.handle.net/10938/23280

Lisker, Leigh & Arthur S. Abramson. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *WORD, 20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Schoonmaker-Gates, Elena. (2015). On voice-onset time as a cue to foreign accent in Spanish: Native and nonnative perceptions. *Hispania, 98*(4), 779–791. https://doi.org/10.1353/hpn.2015.0110

Stölten, Katrin, Niclas Abrahamsson, & Kenneth Hyltenstam. (2015). Effects of age and speaking rate on voice onset time: The production of voiceless stops by near-native L2 speakers. *Studies in Second Language Acquisition, 37*(1), 71–100. https://doi.org/10.1017/s0272263114000151

Appendix

Questionnaire

Where are you from?

How old are you?

What is your gender?

What is your ethnicity?

Did you grow up in a home in which Spanish was spoken?

How long have you been learning Spanish?

How much of that time was in school?

Was any of that time spent in an immersive environment such as a mission or study abroad? If so, how much?

Interview (English)

What was your motivation for learning Spanish?

How do you feel about Spanish as compared to English?

Give a rating for both Spanish and English on a scale of 1 to 10 (1 being the least cool and 10 being the most cool).

Were you originally excited about learning it, and are you still excited now? How do you feel when you hear Spanish being spoken?

How do you feel when you speak Spanish?

Story Passage (Spanish)

Érase una vez que había una mamá cerda que tenía tres cerditos. Ella los amaba mucho, pero no había suficiente comida para alimentarlos, así que los cerditos tuvieron que ir a buscar su suerte.

El primer cerdito decidió ir al sur. Encontró a un granjero en el camino que estaba llevando un atado de paja. El cerdito preguntó respetuosamente: "¿Podría por favor darme esa paja, para que yo pueda construir una casa?". Como el cerdito dijo "por favor", el granjero le dio la paja y el cerdito construyó una bella casa. La casa tenía paredes hechas de paja, un piso hecho de paja, y adentro... una cómoda cama hecha de paja. Después de construir la

casa, el cerdito decidió tomar una siesta en su cama hecha de paja. De pronto, el gran lobo malo llegó y olió al cerdito dentro de la casa, y su boca comenzó a hacerse agua. "¡Mmmmm... emparedados de tocino!" Así que el lobo tocó la puerta de la casa hecha de paja y dijo: "¡Cerdito! ¡Cerdito! ¡Abre la puerta!"

Story passage was extracted from https://www.thespanishexperiment.com/stories/threepigs