Claire Benét Post
ANTH 202
Archaeological Case Study
Word Count: 1997

Colonization's Influence on Quechua Glottochronology

Language change is as inevitable as the turn of the seasons; it is bound by the same flux that changes species over time and the drives of humanity to create new things to talk about (Aitchison 2001, Trask 2013). Despite what older generations love to admonish as the deterioration of language as new generations acquire their mother tongue, it would be strange if language were a process immune to the moving power of change. Yet, in the case of Quechua, a language spoken in Peru, Bolivia, and Ecuador by 8 million speakers today, the language has not simply changed due to the natural phenomenon of time but has been warped and linguistically modulated by Spanish (Américo 2017). Thus, an important question arises of how exactly Quechua has changed over time due to colonization.

There are several typical routes of language change. Language learning, wherein language is changed as is it, is imperfectly transmitted from one generation to the next through regularization (Liberman 2003). Most integrally to this paper, language contact impacts language change through migration, trade, and colonization. To elucidate this process, I will produce a glottochronology of how the colonization of the Spanish has affected Quechua speakers in Peru in terms of lexical change. In general, the impact of colonization on a colonized language is a difficult topic to study because colonization is inherently one group dominating and rewriting the history of another. Nevertheless, this makes the discovery of the buried linguistic history of the indigenous peoples of the Americas even more critical in recognizing the importance of truth for these groups and reconciliation of the past and in creating steps for future language preservation.

Lexical change is the change in the meaning or use of a word which is also known as polysemy (Warren 2013). This is the generational shift of language users to prefer a certain lexicon over another (Warren 2013). Lexical change is amongst the most common forms of lexical change and is incredibly susceptible to 'catastrophic events' such as colonization and conquest that alter the prestige norms (Li 2013, 1). First, "aliens in the lexicon," or loan words, appear marked as code-switching; the foreign word then becomes borrowed and eventually is relexified into the local layer (Li 2013, 1).

Before the first interaction between Spanish conquistadors and the people of the Incan Empire, the languages of what is now Peru, Bolivia, Columbia, Ecuador, Brazil, and Chile were rich in linguistic diversity and completely separate from Spanish (Von Gleich 1994, 80-81). The two main language families present in Peru included the Quechua language family and the Aymara language family (Adelaar 2017, 69). Reports to Spain from Cieza de León in 1550 and Garcilaso de la Vega in 1609 suggested that there between hundreds and thousands of languages and dialects between these families (Von Gleich 1994, 81). Quechua particular to Cuzco was

considered the hierarchic pinnacle by the Spanish crown once they arrived, but in reality, the Chincay southern variety was spread as the "general language" to the Incans in addition to Aru, Aymara, Puqina, and Mochica (Von Gleitch 1994, 80-81). Quechua saw impressive growth of the language due to its policy decisions during the Inca expansion period of 1430-1532 that necessitated linguistic policy to maintain control of conquered territories (Von Gleitch 1994, 82). Although, unlike Spanish linguistic policy to come in the following decade, Incan policy was by no means assimilationist but promoted the Chinchay variety as the *lengua general*, because it was the mostly widely understood (Von Gleitch 1994, 82-83; Garcilaso de la Vega 1609; De León 2005, anonymous Jesuit 1968 [1594]).

$$lev_{a,b}(i,j) = \begin{cases} max(i,j) & \text{if min(i,j)} = 0, \\ min \begin{cases} lev_{a,b}(i-1,j)+1 \\ lev_{a,b}(i,j-1)+1 \\ lev_{a,b}(i-1,j-1)+1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases}$$

Figure 2: Levenshtein Distance. Source: (Majumdar 2022)

From this evidence, it can be presumed that some lexical change occurred pre-colonization to spread the Chinchay variety of Quechua throughout the Andes. This is also supported by overlaps seen in vocabularies today between the different Amerindian language varieties. I obtained a 25,000 word dictionary spreadsheet from 2006 from a site that works to provide translations for Quechua to many other languages—including Indo-European and Amerindian languages. When 'Quechua' is mentioned, I am referring to Quechua written in Southern Quechuan orthography—which is a standard combining features of Cusco and Ayacucho Quechua and uses the present official writing standard of Qusqu-Qullaw in Peru that is provided as the main dialect within the Quechua dataset (Runasimi.de 2006). To see the similarity between Quechua and languages within a similar language family, as well to see the similarities between Quechua and its colonizer's language, Spanish, I calculated the Levenshtein distance score between words (Majumdar 2022, Zach 2020). The Levenshtein distance[1] is a standard glottochronological formula that finds the minimal number character changes between words and is used to decipher lexical change over time (Serva 2008; Haldar 2011; Yujian 2007).

[1] A Levenshtein distance of 0 means no change is needed to make the words the same, thus they are the same word. A Levenshtein distance of 1 means one character change is needed between words: "abril" and "April." The higher the lexical similarity score, the more different the two words are. A general intuition is that once the score is over 6 or 7, the words are entirely different.

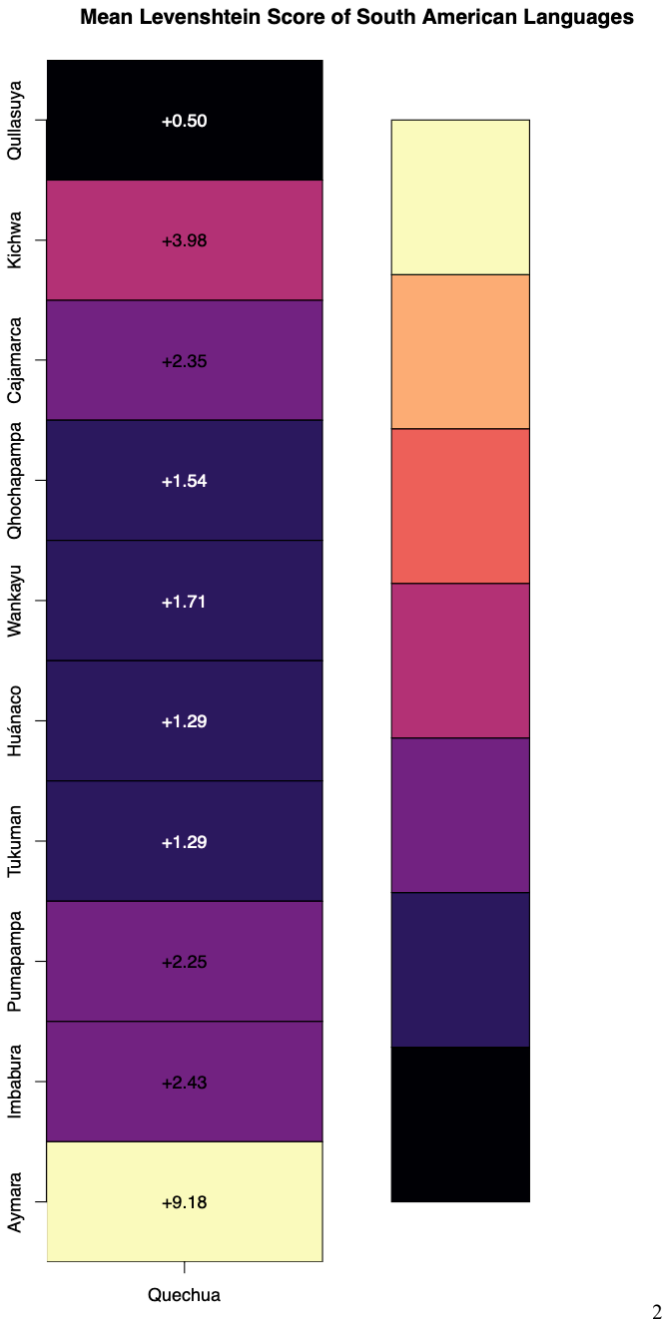


Figure 1: Quechua Similarity to Other Amerindian Languages. Source: (Post 2022).

[2] Qullasuyu is of Cusco-Boliviano and is included from a list of words from Qusqu and Bolivia, written in official Bolivian standard of Southern Quechua, i.e. with a, i, u and j. Kichwa of the Shukllachishka Kichwa type of Quichua Unificado del Ecuador is written in Unified Kichwa standard of Ecuador. Cajamarca is spoken in Peru, written with a, e, i, o, u and mb, ng, nd. Qhochapampa, which also known as Cochabamba, uses former Bolivian writing system with a, e, i, o, u. Wanka is known as Wankayu and is spoken in Junín, Peru. Huánuco is spoken in Huallaga, Peru, written in David Weber's Hispanic writing system. Pumapampa is a variety of Ankash is spoken in Ancash, Peru: Pomabamba, written with a, i, u, ay, aw, uy (Runasimi.de 2006). A note is that this data is not totally complete and does not include all translations between 'Runa Simi.'

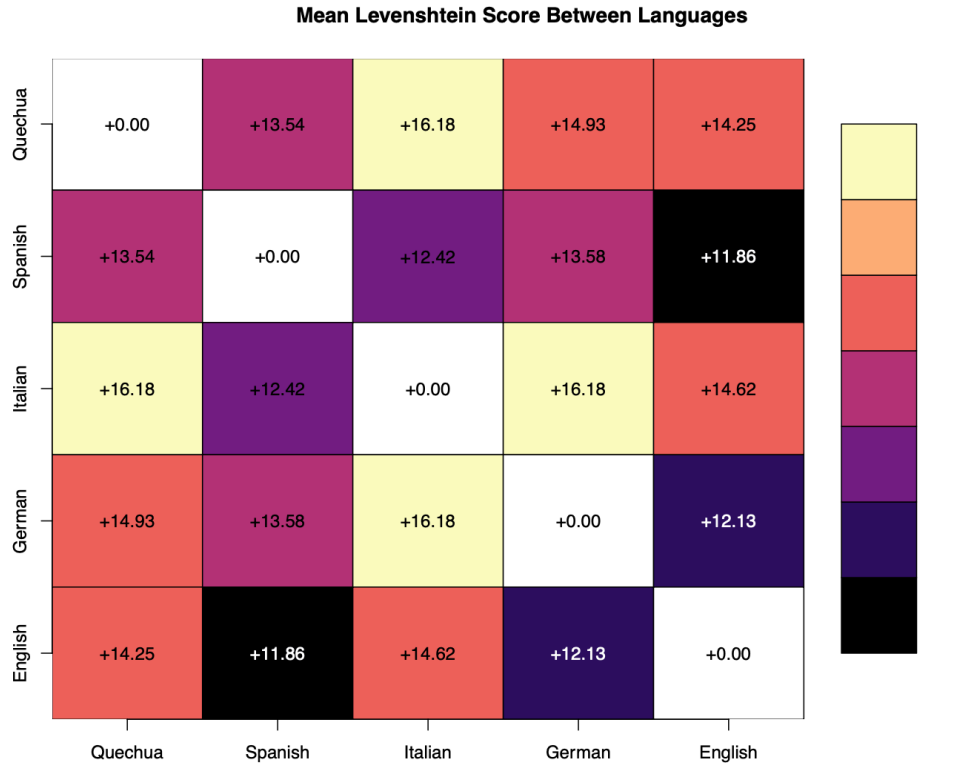
Viewing Figure 2, one may see Quechua's similarity to other languages that originated in South America. Qullasuya, Cajamarca, Qhochapampa, Imbaburu, Tukuman, Wankayu, Huánaco, and Pumapampa are dialects of Quechua spoken in different Peruvian districts and thus are very similar to the general Quechua. Kichwa is closely related to Quechua and is spoken in Bolivia. Aymara, with more than a million speakers currently, differs the most from Quechuan as it stems from a different protolanguage: Quechua stems from proto-Quechua and Aymara stems from proto-Jaqi (Proulx 1987, 91). This is reflected in a mean Levenshtein distance more like those between other divergent stem languages such as English, German, and Spanish respectively.

Post the arrival of the Spanish, major language shifts began to occur with the two often opposing goals of Christianization of the Incan people and holding administrative power over them (Von Gleitch 1994, 83; Durston 2007). Initially, Pizzaro in 1527 and other administrators used Quechua as the *lingua franca* throughout the Peruvian region due to the lack of bilingual speakers of Spanish (Von Gleitch 1994, 84). The regulations surrounding Castilianization policy tended to be revoked or changed within period of last then ten years, which it can be argued the inconsistency of such language policies helped to preserve the Indigenous languages of the region (Von Gleitch 1994, 84). Policy largely fluctuated on how best to evangelize the Amerindians to the Christian faith. In 1634, Philip IV's order to use "the gentlest possible means" to teach the people Spanish was countered by Council of the Indies assertion that cosmic vision provided in the indigenous languages of the Americas could not convey the same cosmic vision as the Christian faith (Von Gleitch 1994, 86).

Despite this, a new type of Quechua began to appear that combined both Spanish and Quechua for the purposes of evangelization called "Pastoral Quechua" (Durston 2007). This language was an institutionalized effort over the 16th century to use translation and codification to produce texts that took advantage of Quechua's unique poetic and grammatical resources to make connections between Christianity and Andean religions (Durston 2007). Peruvian clergy and indigenous assistants disseminated the Pastoral Quechua through liturgy, sermon, and church performance that was specific to combing the language of the partitioners through codification (Durston 2007). Some instances of grammatical changes that stemmed from Pastoral Quechua included the use of the Spanish nominal plural '-s' over the Quechua suffix /-kuna/ in some Bolivian varieties (Durston 2007, 182). Quechua has a complex system of evidentiality markers that were also poorly understood by Spanish missionaries or linguists alike as the /-mi/ evidentiality marker was claimed to "adorn" a sentence—"adorno mucho la tal oración" (de Santo Tomás 1995 [1560], 133; Durston 2007, 183; Faller 2002). Santo Tomás, a linguist who was an earlier writer of Quechua orthographic features in 1560, also misunderstood the phonetics of Quechua, believing the two systems to be made very similar by God, and instituted a writing system that did not distinguish between the phonetic /k/ and /q/ and misrepresented many vowels (Durston 2007, 189). The teachings and standardization of the Quechua writing system undoubtably had an effect foremostly on a previously unwritten language that was redescribed in terms of incomplete Spanish phonetics. In all, the evangelization of the Andeans by the Spanish

crown introduced lexical changes particular to the Christian cosmovision and had major impacts on the fledging writing system that was instantiated.

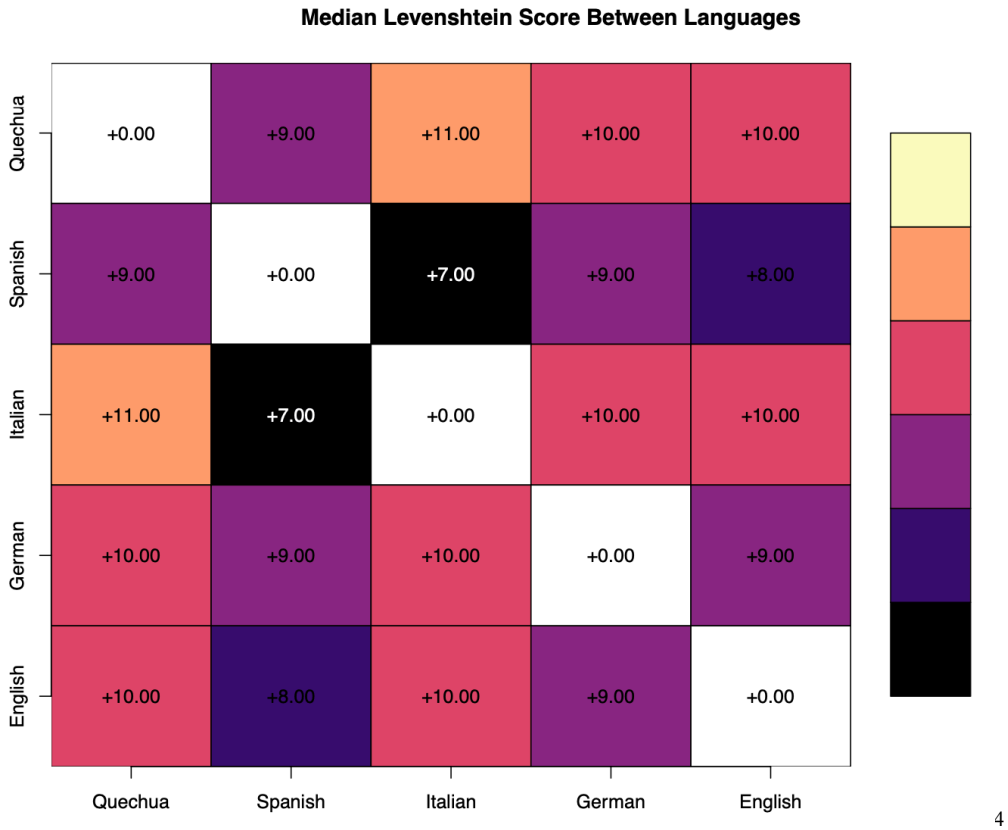
The period between 1600 to 1700 saw similar patterns as colonization continued with notable additions to the lexicon related to class systems and the *encomienda* system. The next largest change in Quechua occurred after Charles III's decree of compulsory Castilianization that was in line with the expulsion of the Jesuits in 1767, who were "zealous defenders of Indian languages and cultures," and the closure of the Quechua chair at San Marcos University in 1783 due to the rebellion of Tupac Amaru (Von Gleitch 1994, 86). The structural economic changes that occurred during this period as the mechanization of farming and mining equipment transpired to increase international exports saw Spanish take hold in lexical spaces were Quechua once held dominance (Von Gleitch 1994, 87).



[3]

Figure 2: Similarity between Quechua and Indo-European Languages. Source: (Post 2022).

[3] I cleaned this data to some extent to get rid of bias in the data that made the scores larger than they should have been. Originally, some definitions would give many different translations. For instance, "chaskichiy" was translated as "to give or reach something; to deliver; to entrust; to give or entrust." So, I shortened it to translate to "to give or reach something." Italian was still biased after cleaning to have longer definitions with extra, unneeded information.



[4]

Figure 3: Median Levenshtein Distance between Indo-European Languages. Source: (Post 2022).

To view the change in lexical similarity due to colonization Figure 2 and Figure 3 display Quechua's similarity to languages outside of its language family. It also shows similarity between the Indo-European languages for a point-of-reference to the differences that should be present. For instance, Spanish and Italian are logically the more similar as they split about 1000 years ago on the language tree compared to Spanish's split from English and German almost 4000 years ago (Serva 2008, 3). One can see that Quechua is the most like Spanish which coheres with the language interaction between the two for the past 500 years.

Certain parts of speech are more susceptible to language change than others: closed-class parts of speech that includes determiners, pronouns, conjunctions, copular verbs, and prepositions are less susceptible, while open-class parts of speech that include nouns, verbs, adjectives, and adverbs are more susceptible (Münte 2001). Industrialization and urbanization have caused language change across the world as new words have to be made to describe the changing world around us—for instance, the noun 'computer' had to be coined to describe a mechanical computing device. This synthesis adds many new words from Spanish and other languages into Quechua over time. It also works in reverse as Quechua has added words to Spanish and English such as "alpaca," "condor," "papa" (potato), and "jerky."

[4] Showing the median is helpful because it mitigates the bias from overly long Italian scores.

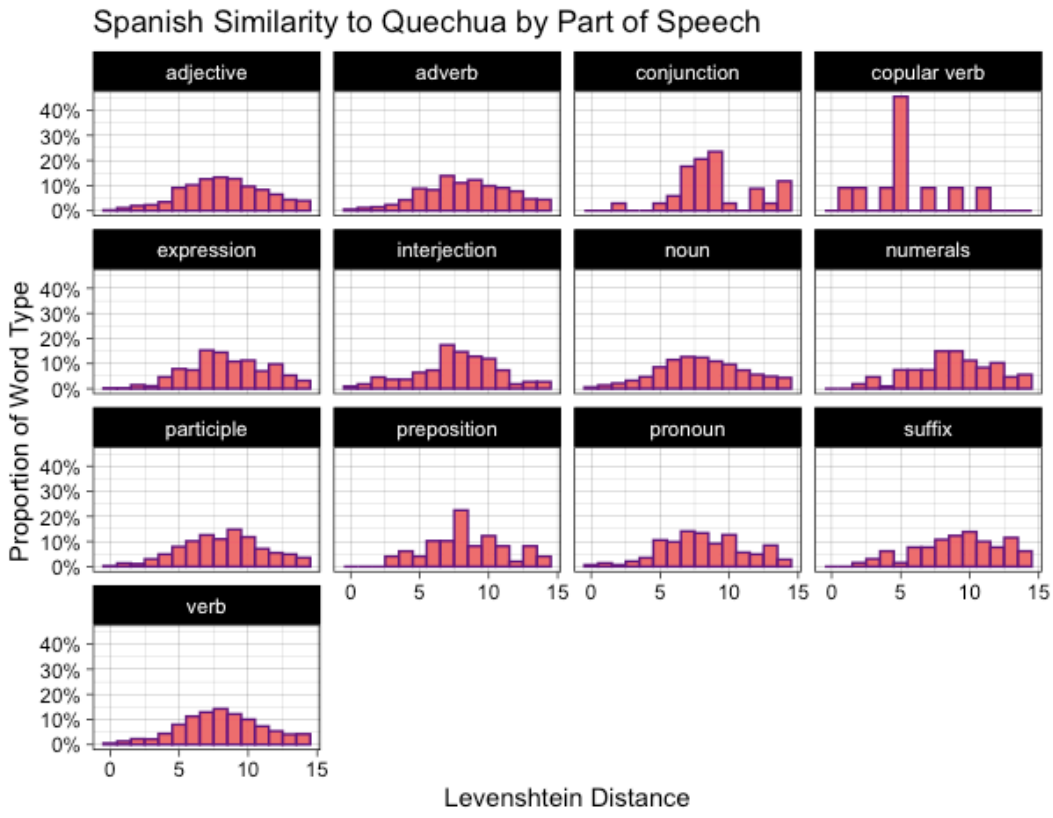


Figure 4: Spanish Similarity to Quechua by Part of Speech. Source: (Post 2022).

My hypothesis when starting data collection was that lexical change in Quechua open-class parts of speech would be more affected by Spanish, but also some closed-class parts of speech may have seen alteration due to the standardizations and relexification processes of Spanish. Figure 4 shows the Levenshtein distance of different parts of speech in Quechua to Spanish. The distributions are fairly normalized across all parts of speech, with no seeming particular bias towards 0, an exact match, for any particular part of speech.

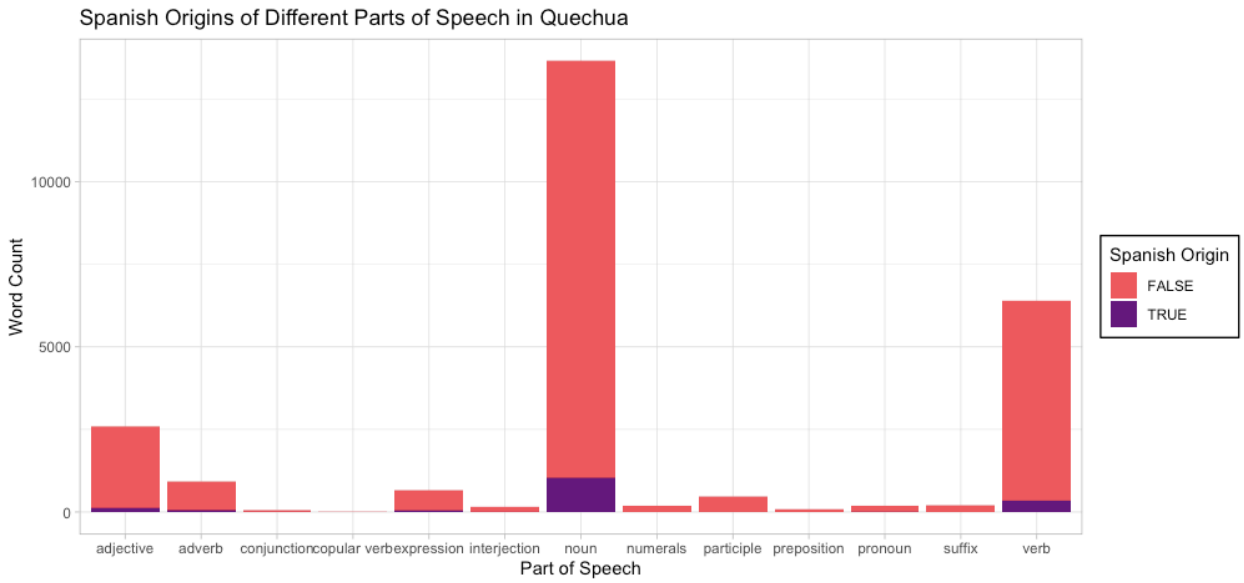


Figure 5: Count of Quechuan words with Spanish origins by part of speech. Source: (Post 2022).

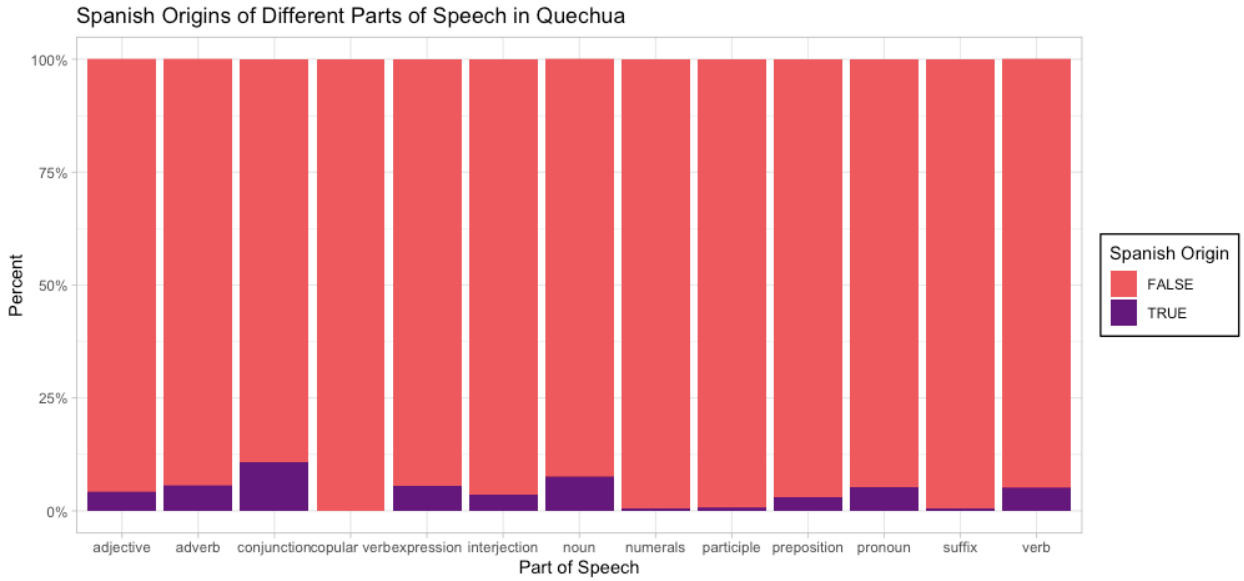


Figure 6: Percent of each part of speech in Quechua with Spanish origins. Source: (Post 2022).

The dataset also provided if words had Spanish origins within the parts of speech that provides more illumination on my query. Figures 5 and 6 display both the word count and percent for each part of speech and their origins. Notice that in Figure 5 there is a reasonably large number of nouns with Spanish origin which is conducive to the open-class theory of language change. Interestingly though, as seen in Figure 6, the conjunctions have been significantly influenced by Spanish. This is an alarming finding because conjunctions are a closed-class part of speech and may indicate Spanish's encroachment on this particular dialect.

In some places such as the Ecuadorian Highlands, Quechua has become a *Media Lengua* of Quechua grammatical structure and nearly 90% Spanish lexicon (Muysken 1981, 52).

Overtime, hundreds of languages and dialects throughout the Andes have disappeared due to the diglossic effects of Spanish colonization. If the data could be found in the form of large dictionaries over time, it would be informative to see how this form of Quechua looks in the past and how it looks today to discern a trajectory of the language into the future. Even still, the preservative actions being taken through policy and documentation is critical to preserving the rich linguistic diversity of the languages of the Andes.

Works Cited

Peer Reviewed:

Adelaar, Willem FH, and Garatea C. Godenzzi JC. "Diversidad lingüística en el Perú precolonial." Literaturas orales y primeros textos coloniales (2017): 67-81.

Aitchison, Jean. Language change: Progress or decay? Cambridge university press, 2001.

Américo Mendoza-Mori. "Quechua Language Programs in the United States: Cultural Hubs for Indigenous Cultures." Chiricú Journal: Latina/o Literatures, Arts, and Cultures, vol. 1, no. 2, 2017, pp. 43–55, https://doi.org/10.2979/chiricu.1.2.05. Accessed 30 Apr. 2022.

Anónimo, Jesuita. "Relación de las costumbres antiguas de los naturales del Pirú." Biblioteca de Autores Españoles, Crónicas Peruanas de Interés Indígena 209 (1968): 161.

De la Vega, Garcilaso. "1609: Comentarios reales de los Incas." Lisbõa: Crasbeek (1963).

De León, Pedro de Cieza. Crónica del Perú: el señorío de los Incas. Vol. 226. Fundación Biblioteca Ayacuch, 2005.

Domingo de Santo Tomás, and Raúl Porras Barrenechea. Lexicón, o, Vocabulario de la lengua general del Perú / por Domingo de Santo Tomás. Ed. facsimilar / publicada, con un prólogo, por Raúl Porras Barrenechea., Edición del Instituto de Historia, 1951 [1560].

Durston, Alan. Pastoral Quechua: The History of Christian Translation in Colonial Peru, 1550-1654. University of Notre Dame Pess, 2007.

Ferguson, Charles A. "Diglossia." word 15.2 (1959): 325-340.

Haldar, Rishin, and Debajyoti Mukhopadhyay. "Levenshtein distance technique in dictionary lookup methods: An improved approach." arXiv preprint arXiv:1101.1232 (2011).

Münte, Thomas F., et al. "Differences in brain potentials to open and closed class words: Class and frequency effects." Neuropsychologia 39.1 (2001): 91-102.

Muysken, Pieter C. "Halfway between Quechua and Spanish: The case for relexification." (1981).

Proulx, Paul. "Quechua and Aymara." *Language Sciences* 9.1 (1987): 91-102.

Rodríguez, Yliana. "Spanish-Guarani diglossia in colonial Paraguay: A language undertaking." The Linguistic Heritage of Colonial Practice 13 (2019): 153.

Serva, Maurizio, and Filippo Petroni. "Indo-European languages tree by Levenshtein distance." EPL (Europhysics Letters) 81.6 (2008): 68005.

Trask, Larry. Language change. Routledge, 2013.

Von Gleich, Utta. "Language spread policy: The case of Quechua in the Andean republics of Bolivia, Ecuador, and Peru." International Journal of the Sociology of language1994.107 (1994): 77-114.

Warren, Beatrice. "Laws of thought, knowledge and lexical change." Historical Semantics and Cognition 13 (2013): 215.

Yujian, Li, and Liu Bo. "A normalized Levenshtein distance metric." IEEE transactions on pattern analysis and machine intelligence 29.6 (2007): 1091-1095.

Unreviewed:

Faller, Martina Theresia. Semantics and pragmatics of evidentials in Cuzco Quechua. Stanford University, 2002.

Li, Chin-An Arnold. Lexical change and variation in Taiwanese literary texts, 1916–1998: A computer-assisted corpus analysis. University of Hawai'i at Manoa, 2000.

Liberman, Mark. “Types of Language Change.” Linguistics 001 -- Language Change and Historical Reconstruction, https://www.ling.upenn.edu/courses/Fall_2003/ling001/language_change.html.

Post, Claire Benét. “Quechua Glottochronology: Analysis of Quechua and Other Indo-European and Mesoamerican Languages.” GitHub, 27 Apr. 2022, https://github.com/clairepost/QuechuaGlottochronology.

Runasimi.de. “Quechua Multilingual Dictionary.” Edited by Philip Jacobs, Runasimi.de: Runasimi (Quechua, Quichua): Ayakuchu, Qosqo (Qusqu), Qhochapampa, Imbabura, Tukuman, Ankash (Waylas), Aug. 2006, https://runasimi.de/runaengl.htm.

Zach. “How to Calculate Levenshtein Distance in R (with Examples).” Statology, 18 Dec. 2020, https://www.statology.org/levenshtein-distance-in-r/.