|  |
| --- |
|  |
| No Room to Breathe: A Corpus Analysis of Senegalese French |
| Tyler Lanni  5/20/2022  Language Contact |

|  |
| --- |
|  |

**Introduction**

The identity of a culture and the language which it speaks is intertwined. This is experienced at different levels of a society: from state and legislative to the general population. In heavily multi-ethnic areas, a common language is employed to facilitate ease of interaction between various ethnic groups. It is commonplace in post-colonial states such as Senegal, the Democratic Republic of Congo, Ivory Coast, and other West African countries to have that common language be French. This extends far beyond the streets and markets, however, and is also institutionalized by those in political control. This paper will analyze Senegal’s unique situation where the French language, originating from a colonial power, was determined to be an *official* language while indigenous languages were labeled as *national* languages.

How countries officially regulate languages is through their language policy. Ideas, laws, rules, and practices determined by official organizations set precedent for how that country’s institutions and populations are expected to communicate. Language policy has an important role in the cultural identity of a country. Claiming a language as an official language has implicit effects through the society. Whether a school may educate with a national language, or the official language has educational effects on its students. Inclusion and representation of a culture’s identity may be rejected by its lack of representation within the media industry. A politician’s comments about which language should be the language of the people may create a sense of unity for some, but isolation for others. These practices and regulations can be seen in formal and informal designations. Both are important in a country’s language identity.

This paper attempts to analyze the effect language policy and public opinion have on the media industry in Senegal. In Senegal, newspapers are heavily influenced by the language policy of their country. All national newspapers are written in French, the offical language of Senegal. Public opinion regarding their historical colonizers has recently caused a number of riots in the capitol, Dakar (Thomas-Johnson, 2021). There is a feeling of inequality and misrepresentation amongst the general population. The French Language is seen as the language of the elites, while other national languages are spoken by the general population. French is only spoken by roughly 10% of the population (Sall, 2009). How does Senegal’s history and language policy affect French in Senegal?

**Previous Works**

Gulinov et al. (2018) represent France’s Language policy in multiple levels. These levels are situated in a vertical sense and each play a role in how effective language practices and regulations are implemented. The levels were summarised as the following:

1. the first persons of the state (the president, ministers), public figures, officials capable of influencing the current language situation in the country, initiating language laws and regulations that determine the state policy in the field of language as a whole;
2. the institutions of language policy – state and public structures, whose activities are aimed at preserving the identity of the language, as well as regulating the processes of the language development;
3. consumers of language reforms developed by state institutions.

The author claims that each level has a part to play in how effective a language policy is. They also posit that “Ideological principles and practical measures in the field of language policy are interdependent and inseparable” (Gulinov et al., 2018). Because linguistic policy is an integral part of national policy, a country cannot make an effective language policy without considering the intricate layout of mixed nationalities within its borders.

How does France’s historically prescriptive behavior translate onto heavily French influenced nations? Is the intricate verticality of France’s language policy and purist beliefs about French applicable to other countries where it is the official language? Michelman (1995) discusses the comparisons between English and French language policies in Africa and their effects on African literature. Michelman relates a recent event in which the French government proposed that foreign words be prohibited from virtually all public government and commercial communications if a suitable equivalent exists in French. This kind of restrictivism, Michelman notes, is pervasive in the culture and language of France (Michelman, 1995)**.** The language is tied to the French identity, anything other than standard French offensive and, by some accounts, punishable. This idea of French purism extended to their colonies. An example of early colonial education can be seen in Michelman’s writing:

‘French only is to be used in schools. It is forbidden for teachers to speak to pupils in the local languages.’ But beyond this prohibition was the intimidating ban imposed on the students themselves by means of the infamous system known as “le symbole”: an object (such as a box of matches) was circulated from student to student as each was caught by his classmate speaking his native tongue. At the end of the day the unlucky holder of the “symbole” was subjected to corporal punishment by the teacher. Moreover, neologisms and the ungrammatical French were severely supressed. (p. 219)

Given that part of France’s language policy includes individuals using the language, as noted by Gulinov et al. (2018), what effect do these ideas have on Senegal French today?

Michelman describes the difference between English and French expansion concerning identity. For the French, colonies were meant to take on French identity, with their language being instrumental in the undertaking. Africans were meant to assimilate into French culture and conform to French ideas, identity, and language. This contrasts the British who desired “differentiation” between British and African subjects. The British envisioned “the separate development of African peoples", maintaining a gap between Europeans and Africans in social and cultural aspects. Ethnocentricity and superiority displayed itself differently between the two colonizers. Where France wanted to spread their self-proclaimed superior identity and culture into the colonies, the British desired Africa to develop separately. This difference in colonial policy extended to how each country dealt with indigenous languages.

English colonies were more inclusive regarding indigenous African languages. Colonies were encouraged to use African languages in education, literature, and even administration. The French, in contrast, desired cultural imperialism. Schools were forced to educate in solely French. Teachers were forbidden to speak to students in their native language. Corporal punishment was encouraged when improper French was spoken by students. Creoles and non-standard French was labeled a language for less-than-human individuals. To this day, countries previously under French control use French in literature, state journalism, and newspapers (Sall, 2009; Michelman, 1995; Mc Laughlin, 2019).

Returning to the idea that language policy is affected by “consumers of language reforms developed by state institutions” (Gulinov et al., 2018), a previous corpus analysis of Greek and German metalinguistic discourse shows the extent to which language is shaped by the public and media. Moschonas & Spitzmüller (2010) analyzed prescriptive discourse, or more broadly somone telling another person they’re misusing language, and deconstructs the correctives into different categories: lexicon, pragmatics, semantics, morpho-syntactic, phraseology, orthography, and miscellaneous. It’s here, in the public’s prescriptivism, we can see language policy verticalities in Germany and Greece, countries other than France. This gives us support to extend Gulinov et al.’s findings to France’s previous colonies like Senegal. The language of the media is set by officials in the government, the media then distributes language to the public, and the public both utilizes the language in their own discourse and critisizes the media’s use of language through prescriptivism.

**Methodology and Theoretical Orientation:**

Using a corpus analysis to compare Standard French from a corpus of French news articles against French from Senegalese newspapers, can we see a notable difference? Any differences discovered will then be analyzed given the context of Senegal’s history, current language policy, and public opinions about the French language.

This corpus analysis will focus on lexical items. We will analyze differences in unique terms, lemma versus word count, and sentence length.

Data was gathered through web-scraping newspaper websites. For France-French the newspaper “Le Monde” was web-scraped. APS (Agence de Presse Sénégalaise) was web-scraped for Senegalese French. 15,000 sentences were added to each corpus before cleaning and pre-processing. The corpus analysis was to be based on 10,000 sentences of France-French versus 10,000 sentences of Senegal-French. The extra 5000 sentences for each provided enough room for deletion of bad data.

After cleaning and removing poor sentence constructions (incomplete sentences gathered due to formatting issues on the website), data was then pre-processed. Sentences were lemmatized and stripped of punctuation for analyzing lexical diversity.

Before pre-processing was completed, however, it should be noted that Le Monde showed preference for distinct puntuation choices. The use of cheverons (“«” and “»”) occurred incredibly frequently in their articles. Senegalese French did not display cheverons at all. This could be the institution’s individual preference or be linked to something broader. More analysis is required.

A picture containing bar chart

Description automatically generatedA picture containing bar chart

Description automatically generated**Results:**

A picture containing chart

Description automatically generatedA picture containing bar chart

Description automatically generated

France-French (FF) had a word count of 153,616; a unique-term count of 32,932; a Type-Token Ratio (TTR) of 0.214379; a Root Type-Token Ratio (RTTR), otherwise known as Guirauds Index of Vocabulary Richness, of 84.023328. Senegal-French (SF) had a word count of 257,158; a unique-term count of 29,583; a TTR of 0.115038; a RTTR of 58.336745.

FF had a higher diversity of words used in its newspaper compared to SF. SF’s corpus contained 103,542 more words, but 3,349 fewer unique tokens. In other words, FF used less words in total but higher lexical diversity. This must mean that the average SF sentence had more words than its FF counterpart and their sentences are likely to be longer than FF sentences.

FF’s RTTR (RTTR = types/√tokens) score of 84.02 is much higher than SF’s of 58.34. This score is useful in determining the lexical diversity of a corpus. The distinction between the total amount of words and which of those are unique varies drastically between FF and SF. As noted previously, SF used more total words but repeated a majority of them. FF was much more likely to introduce a new word to the corpus than SF.

The data also held information regarding words occurring just once (1-time words) and lemmas which occurred just once (1-time stems). The ratio of all 1-time words per 1-time stems (1-time words / 1-time stems) was higher for FF than SF. FF had a ratio of 1.81 while SF had a ratio of 1.52. This shows the FF corpus containing more diversity amongst word construction. A lemma in the FF corpus was more likely to be used in constructing various words, where as SF was more likely to contain a lemma with just one word construction. As an example, the lemma <prod> can be used to construct the words <production>, <produce>, <product>, etc. FF was more likely to have a diversity amongst words but also A picture containing chart

Description automatically generatedA picture containing chart

Description automatically generatedwas more expressive when utilizing lemmas.

Following 1-time words and their distribution in both corpora, we can see a trend in words occuring from 1-5 times. FF had a much higher count of 1-time, 2-time, and 3-time words eventually converging with SF in the count of 5-time words. The trend seems to be the same past 5-time words, but this does show, again, that FF was more willing to employ unique words compared to SF. This same trend occurs when analyzing the stems used in each corpus, but to a lesser extent.

**Discussion:**

The lexical diversity of each corpus shows the difference between two situations of a country using French. One, FF, holds French as its official and native language. SF has French as an official language but not a native one. It is clear from the results that FF has a higher lexical diversity, but it stretches beyond words. This study shows FF’s expressiveness in lemma diversity as well as word diversity. What does this mean for SF compared to FF though?

One of the key characteristics of a pidgin or creole (P/C) is that the lexicality of the language becomes simplified. Siegel (2008) describes the simplicity as quantitative rather than psycholinguistic, as the *simplified* language is no easier to process than a more *complex* language. When calling a P/C simplified, Siegel (2008) does claim that there is a reduction in a P/C’s lexicon compared to its lexifier language.

One other point of interest which may provide evidence of pidginization is our word-to-lemma ratio. FF had a higher percentage of word constructions per lemma. This displays the FF’s more diverse morphology compared to SF, allowing the argument that SF is morphologically simpler than FF according to Siegel (2008). With a simplified lexicon and morphology, is there enough evidence to say that SF is undergoing pidginization, or is there another reason for SF’s simplification?

Siegel (2008) states that simplification is not brought on by the creation of P/C’s, however, and determines Second Language Acquisition (SLA) to be the cause for simplified features. In other words, pidgins and creoles are a result of SLA’s simplification, not the cause. That’s not to say that pidginization doesn’t entail a quantifiable sense of simplification. Even amongst authors who challenge the overall simplicity claim like Ansaldo (2017) and Mufwene (2013), there seems to be an acceptance that P/C’s display a quantifiable simplicity in their morphologies. More accurately, there is a less diverse surface-level morphology.

With the acceptance of simplified lexicality and morphology, what can we say about SF compared to FF? There isn’t enough evidence within this study to determine if SF is in a stage of pidginization or creolization. Further analysis on features other than lexicality and morphology can aid in that determination, but we can use our current data to see that SF shows effects of SLA.

The viewpoint that SF shows simplified features because of SLA may limit our findings to a few causes: the Senegalese media is simplified because of the writers’ non-native French abilities, the media is simplified for the general Senegalese population, or Senegalese prefer to express themselves and their ideas in other languages, making French less *alive*.

All three possible causes of simplification are likely related to Senegal’s language policy and current education situation. Sall (2009) characterizes Senegal’s language policy into two points:

1. maintain the French language as official language of international communication and

2. promoting national languages into the languages of culture and languages of instruction.

Around 30 languages are spoken within Senegal with Wolof being the major national language. During early education, French is the medium of instruction but there is also a de facto practice of using national languages to bridge learning (Hyter, n.d.). Wolof, being used and understood by over 90% of the population, trumps French’s 10% despite being the official language of communication and education (Hyter, n.d.; Sall, 2009).

There seems to be a strict setting in which French is used instead of a national language. Public opinion is that “French should be used in state institutions, public areas, official speeches, in schools and even unversities” (Sall, 2009), but Wolof is spoken in all environments. Wolof appears to be the national language of choice in Senegal. With Wolof, Senegalese express themselves and their identities. It is even emerging in official speeches, a space which used to be controlled solely by French (Sall, 2009).

There is a growing displeasure with modern French colonialism within Senegal (Thomas-Johnson, 2021). French businesses and the French language are currently seen as items of inequality – remnants of Senegal’s colonial period. It doesn’t help that Senegal’s official currency is still the CFA franc, the same currency used by 14 previous west-African colonies (Thomas-Johnson, 2021). Senegal seems to be pushing away from France and their tongue. The Senegalese want to keep their identity separate from their previous colonizers and are slowly strenghtening their national languages to challenge French as the official language. Wolof is appearing in previously French occupied spaces and is understood by a majority of the population.

Returning to our three possible causes of simplification in SF, we might now have a better idea of which ones are more probable. French is still the official language of the government, media, and education. It is unlikely, though possible, that the simplification is due to the media’s weak grasp of the French language. There isn’t much evidence within this paper or previous literature which would imply that persons in the media have a weak grasp of French, so much as to be measured statistically. This lack of evidence lends more plausibility to social factors, like general opinion about France and the French language, affecting SF’s simplification.

Our second potential cause, one which states that simplification is designed for the general population to ease understanding, might have more weight. Still, though, our analysis does not explicitly describe SF as something that is easier to understand. SF’s less diverse lexicality and morphology does not provide evidence that it is easier for the 90% of Senegalese who don’t speak French to understand. Even if that were the case, recalling that SF had longer sentences on average than FF seems counter-intuitive to facilitation of understanding.

Potential cause number three, Senegalese prefer to express themselves and their ideas in other languages, making French less *alive*, seems to be the most supported option. The revival of languages like Wolof and their increasing presence in official spheres can be related to Senegalese wanting to solidify their African cultural authenticity (Sall, 2009). The Senegalese wish to be addressed officially in a language they fully understand, and French is clearly not that. What this might mean is Wolof being used in all settings other than the typical official settings. Topics and areas of interest discussed in SF might only be used to describe international affairs, not everyday life for example. Wolof is a truer representation of Senegalese culture, and the citizens might feel more comfortable expressing themselves with it.

This idea also resonates with the retelling of Edmond Laforest’s, a prominent member of the Haitian literacy movement, suicide in 1915. A man who “stood upon a bridge, calmly tied a Larousse dictionary around his neck, then proceeded to leap to his death by drowning” (Michelman, 1995). Dramatic as a writer can be, this suicide displays the “dilemma of the non-European writer trapped in the language of the colonizer (past or present)” (Michelman, 1995). France did not just push the French language onto African colonies but their identities, smothering the pre-existing African cultural authenticity.

French may be the official language of Senegal, but it seems clear that national languages, like Wolof, are the breathing languages of its culture. What topics might be expressed in Wolof in Senegal are surely discussed in French in France. SF is most likely used to convey information important to the government – official information meant to be used for international communications, as stated in their language policy. It doesn’t seem a stretch to claim that SF’s simplified lexicality is likely caused by a less diverse topic pool in the media.

**Conclusion:**

Colonialism in Africa left a lasting effect in all countries involved. Though colonialism is deemed a thing of the past, countries like Senegal still see their history affect modern life. Though France attempted to spread French culture throughout their colonies and did so by various means, including language, the main aim was to strip Africans of their identities. Strict, restrictive language policies in the colonies saw indigenous languages sidelined and rejected in official spaces. This was just one tactic in the overall strategy to dominate Africa through de-humanization (Mngomezulu, 2015).

In Senegal, where the official language is French despite only 10% of the population speaking it, we can still see the control France has over the country. Referring to Gulinov et al.’s underlying tertiary system of language policy, we can see the lower level of the pyramid applying pressure to the higher tiers. General opinion about the French language is souring within the country and national languages, like Wolof, are being chosen in place of French. French is losing its control and position as the sole official language. The pressure from the population is pushing up, albeit slowly, against policy makers and laws claiming that French is the official language of Senegal (Sall, 2009).

This rejection of French by Senegalese can be seen in the limited number of spaces that French controls. The lack of SF lexical diversity is likely due to a minimal pool of topics in Senegalese French media. French is used by elites in government, media, and education: laws are passed in French, media disseminates official information in French, and students are educated in French. These official spaces are restricted to most of the population by a language barrier, rejecting the citizens they’re meant to serve. Wolof is used by most of the population to express Senegalese thoughts and lived experiences, attaching itself to the identity of Senegalese. Wolof is seeing an expansion of its influence in all domains of Senegalese life, even those historically restricted to French control (Sall 2009).

**Works Cited**

Ansaldo. (2017). Creole complexity in sociolinguistic perspective. *Language Sciences (Oxford)*, *60*, 26–35. <https://doi.org/10.1016/j.langsci.2016.10.008>

Çöltekin, Ç., & Rama, T. (2022). What do complexity measures measure? Correlating and validating corpus-based measures of morphological complexity. *arXiv preprint arXiv:2204.05056*.

Hyter, Y. D. (n.d.). Cultural connections: From Senegal and West Africa to your classroom. Project Themes: Language Policy & Literacy. Retrieved May 17, 2022, from https://www.cultureconnections.org/themes/language-policy-literacy.html#:~:text=In%20Senegal%20%2C%20the%20language%20policy,cities%20(Dakar%20for%20example).

Mc Laughlin, F. (2019, July 24). Geopolitics of French in francophone Sub-Saharan Africa: Attitudes, language use, and identities. by Ibrahima Diallo. OUP Academic. Retrieved May 20, 2022, from https://academic.oup.com/fs/article-abstract/73/4/650/5537501?redirectedFrom=fulltext

Michelman, F. (1995). French and British Colonial Language Policies: A Comparative View of Their Impact on African Literature. *Research in African Literatures*, *26*(4), 216–225. http://www.jstor.org/stable/3820239

Mngomezulu, M. (2015). Citizenship in Colonial Africa - Journal of Political Inquiry. Retrieved May 20, 2022, from <http://www.jpinyu.com/wp-content/uploads/2015/12/Submission1.pdf?bcsi-ac-1890e3206a556864=2791AF9A00000002C4sJwOENkVvWNKQKaivWaTtr7+VIMwAAAgAAAMq2wACEAwAAAQAAAMrjKQA=>

Moschonas, Spiros & Spitzmüller, Jürgen. (2010). Metalinguistic discourse in and about the media: some recent trends in Greek and German prescriptivism.

Mufwene, S. S. (2013). Simplicity and Complexity in Creoles and Pidgins: What’s the Metric?, Journal of Language Contact, 6(1), 161-179. doi: https://doi.org/10.1163/19552629-006001005

Naida, M. (2016). Senegal's Language Problem: A Discourse of Disparity.

Sall. (2009). Multilinguism, linguistic policy, and endangered languages in Senegal. *Journal of Multicultural Discourses*, *4*(3), 313–330.<https://doi.org/10.1080/17447140903395001>

Siegel. (2008). Pidgins/Creoles, and Second Language Acquisition. In *The Handbook of Pidgin and Creole Studies* (pp. 187–218). Wiley‐Blackwell. https://doi.org/10.1002/9781444305982.ch8

Thomas-Johnson, A. (2021, March 13). *Why is anti-French sentiment spiking in Senegal protests?* News | Al Jazeera. Retrieved May 17, 2022, from https://www.aljazeera.com/news/2021/3/12/senegal-anti-french-sentiments-on-the-rise-amid-ongoing-protests

*WordCruncher: Phrase compare report (N-grams)*. (n.d.). Retrieved May 18, 2022, from http://beta.wordcruncher.com/docs/phrase-compare

Yu. Gulinov D, R. Zheltukhina M, A. Shestak L, G. Slyshkin G, V. Katermina V, D. Chervyakova L. Modern Language Policy: Specifics Of Formation And Development Of The French Vertical. MJLTM. 2018; 8 (2) :160-174