

The Effects of Language Policy on Language Identity in Serbia and Croatia

A thesis presented
by
Adam V. Aleksic

Presented to
The Department of Government
and
The Department of Linguistics
in partial fulfillment of the requirements
for the degree with honors
of Bachelor of Arts

Harvard College
March 2023

Table of Contents

Introduction.....	3
Chapter One: Literature Review.....	10
Chapter Two: Analysis.....	25
Methods.....	25
Sentiment Differences by Language.....	33
Endonyms and Exonyms.....	35
Attitudes Toward a Unified Language.....	40
Overlap.....	43
Conclusion.....	44
Chapter Three: Field Research.....	46
Methods and Street Interviews.....	47
Structure of Language Institutes.....	52
National Narratives.....	59
Salience of Identity.....	64
Retrospective.....	68
Conclusion.....	70
Appendices.....	65
Appendix A: Sentiment Analysis Code.....	75
Appendix B: GPT-3 Prompts.....	78
Appendix C: Sample Interview Guide.....	85
Bibliography.....	86

Tables and Figures

Figure 1.1: Differences in former “Serbo-Croatian” dialects.....	5
Table 2.1: Number of tweets about each language name.....	29
Table 2.2: Number of tweets normalized for population.....	30
Table 2.3: Proportion of discussion about each language name.....	31
Table 2.4: Sentiment differences for tweets about <i>hrvatski jezik</i>	33
Table 2.5: Sentiment differences for tweets about <i>srpski jezik</i>	34
Figure 2.1: Sentiment differences for endonyms.....	36
Table 2.6: Highlighting extreme sentiments in endonyms.....	37
Figure 2.2: Sentiment differences for exonyms.....	38
Table 2.7: Highlighting positive sentiments in exonyms.....	39
Table 2.8: Sentiment differences for tweets about <i>naš jezik</i>	40
Table 2.9: Highlighting extreme sentiments in ambonyms.....	41
Table 2.10: Sentiment differences for tweets with overlap.....	43
Table 3.1: Responses for the name of the language.....	48
Table 3.2: Responses for the difference between Serbian and Croatian.....	49
Table 3.3: Responses for thoughts on <i>naš jezik</i>	50

Introduction

Among linguists, there is a famous adage that “a language is a dialect with an army and a navy.” The boundary between the two can be very nebulous and difficult to define, because all related languages lie somewhere on a gradient of similarity. Drawing the line often comes down to political and cultural factors. For example, Spanish and Portuguese have a lexical similarity of 90% but are considered separate languages, while Cantonese and Mandarin are completely unintelligible but are commonly classified as both being dialects of Chinese.

It’s not a coincidence that the Spanish and Portuguese languages both have their own “army and navy,” while Cantonese does not. Language can be wielded by governments as a tool for both division and unity. The perception of sharing a language with others creates a sense of commonality, while having a separate language from others can serve to alienate them.

The ambiguity inherent to language dichotomy is particularly relevant in Serbia, Bosnia, Croatia, and Montenegro, which until the dissolution of Yugoslavia were all classified as speaking the “Serbo-Croatian” language. Same army, same navy. After they all became their own countries, they respectively began referring to their respective languages as Serbian, Bosnian, Croatian, and Montenegrin. Different armies, different navies.

Linguistically classifying the languages in the former Yugoslav space is immensely difficult because of the distinction between standard languages and dialects. The “standard language” of a country is the literary language used by the government, media, and on official signs. All four countries use the Štokavian

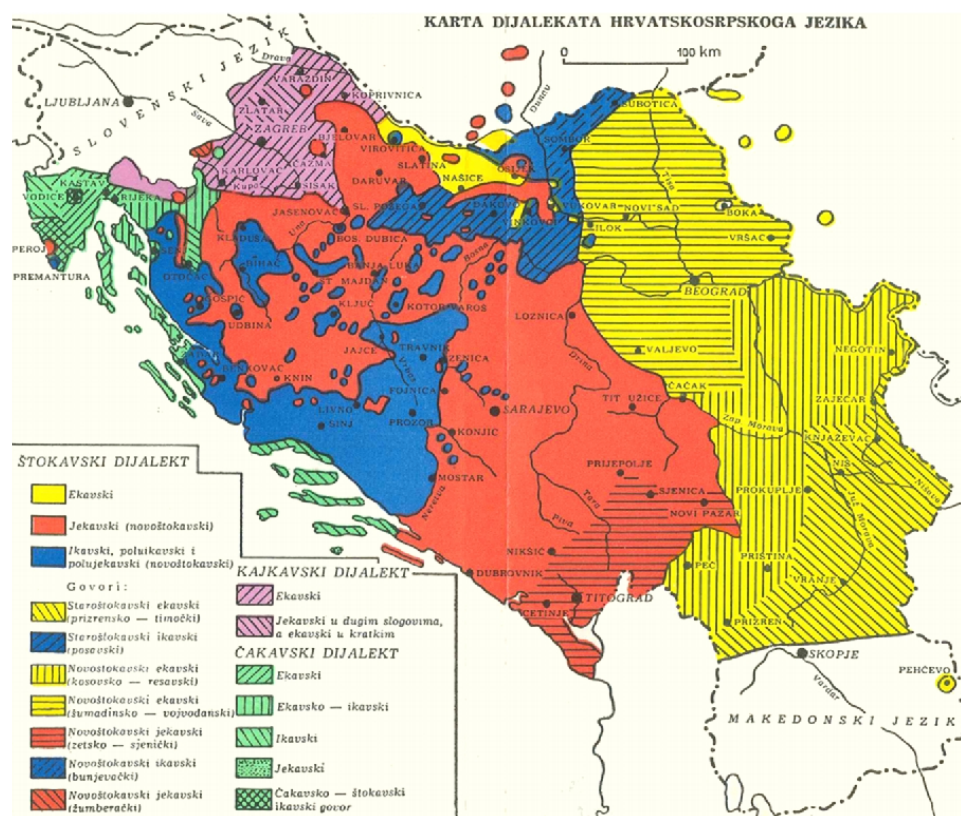
dialect (named after its word for “what,” *što* or *šta*) as the basis for their literary language. However, other dialects, such as Kajkavian in central Croatia (which uses *kaj* as its interrogative pronoun) Čakavian in the historical regions of Istria and Dalmatia (which uses *ča*), and Torlakian in southern Serbia (which contains more similarities to Bulgarian) also exist. Paradoxically, the dialects may differ more from the standard language of their own country than the standard languages do between the countries.

To complicate matters further, each of these dialects can be categorized into accents based on the evolution of the old Slavic phoneme *jat*: Ikavian, Ijekavian, or Ekavian. Serbia uses Štokavian Ekavian as the basis for its literary language, while Croatia uses Štokavian Ijekavian, but individual dialects and idiolects within those countries can take on any permutation of those categorizations.¹ Croatian and Bosnian also only use a Latin-based script, while Serbia and Montenegro use both that and Cyrillic.

The following map is from a 1970 paper by Brabec, Hraste, and Žiković. It attempts to visualize the different dialects across the region:

¹ Stojiljković, Vukašin. “Negotiating the Yat Border(s).” *Wiener Linguistische Gazette*, no. 81, 2017, p. 47.

Fig. 1.1: Differences in dialects across countries formerly categorized as speaking “Serbo-Croatian”



As you can see, there is a practical potpourri of dialects in the former Yugoslav space. Purely from a linguistic perspective, they do not differ enough to be categorized as a language, due to mutual intelligibility and the fact that the structural markers of the languages’ grammars are identical. In 2017, a group of linguists promoting this concept from all four countries drafted the *Declaration on the Common Language*, a document stating that Bosnia, Croatia, Serbia, and Montenegro all share one polycentric language. Over 200 prominent public figures and 10,000 people across the region signed the declaration, sparking new discussion on Balkan language nationalism.

This, of course, relies on the assumption that the definition of “language” should rely entirely on a linguistic definition. Much of the arguments separating Croatian from Serbian, for instance, are cultural or historic. After Serbian linguist Vuk Karadžić created the basis for a standard Serbian language in the early nineteenth century, a group of Croatian intellectuals known as the Illyrian movement and led by linguist Ljudevit Gaj similarly created a standard Croatian literary language, which many Croatian nationalists draw on today as historical proof of a separate language. Counterintuitively, Gaj and the Illyrians were largely in agreement with Karadžić, aiming to create a Štokavian-based language to unify south Slavs. The idea that they “created Croatian” somewhat redefines their legacy. However, in confluence with the different religion, cultural heritage, and Western European influences present in Croatia, there is a basis for an identity-based argument justifying the existence of separate languages.

When the real linguistic relationship between language and dialect is as blurred and meaningless as it has become in the former Yugoslav space, comparing how people think about their language is an important way to understand identity. People in Croatia, regardless of whether they use Čakavian, Kajkavian, or Štokavian, largely identify as speaking “Croatian,” while people in Serbia, regardless of whether they use Štokavian, Torlakian, or something else, mostly say they speak “Serbian.” There are also some people, such as the signatories of the *Declaration on the Common Language*, who are dissatisfied with nationalist efforts to separate the countries’ languages and still identify as

speakers of “Serbo-Croatian” or *naški*, which means “our language” and is an umbrella term encompassing the languages.

Of course, there are many factors that drive the development of language identity. However, Serbia and Croatia are particularly good models for comparing how identity is affected by language policy because, following the dissolution of Yugoslavia, Croatia most rapidly moved to change its literary language from the Yugoslav standard while Serbia did not make any major changes. This thesis explores whether the policy differences between the two countries have impacted how people think about their own language. More specifically, I will attempt to answer the question “how do the different language policies in Serbia and Croatia affect linguistic identity?”

In the first substantive chapter of this thesis, I examine the existing literature on the topic in order to create a framework to contextualize my research. I also use it to further explain the linguistic background of Serbian and Croatian and build a picture of what language identity looks like in the two countries. Finally, I use gaps in the existing literature to justify the relevance of this thesis.

In my second chapter, I perform a sentiment analysis of how Twitter users in Serbia and Croatia are talking about language to find differences in language identity. The function of this is to identify which differences exist so as to understand the current state of identity in the countries and better address the impact of language policy on identity in my third chapter. Based on the social media findings, I find that A) Croats are more likely to exhibit polarized feelings, both positively and negatively, about their own language than Serbs; B) Croats are

more likely than Serbs to view the Croatian and Serbian languages as separate; and C) Croats are more likely to believe that Serbian is a language than Serbs are to believe that Croatian is a language.

In my third chapter, I begin by using interviews with randomly sampled citizens to corroborate the findings from chapter two. I then hypothesize a series of three policy-based mechanisms that explain the differences in language identity and I use interviews with both the random citizens and key policymakers to elaborate on how these mechanisms operate and how they might shape identity. I conclude that the ways in which language policies differently affect linguistic identity in Serbia and Croatia are through A) the structure of language institutes, B) different national narratives, and C) a stronger salience of identity in Croatia.

Due to the complicated relationship between language and dialect in Serbia and Croatia, I find it necessary to elaborate on my word choice throughout this paper. When comparing the policy differences between “Serbian” and “Croatian,” the “Serbian language” refers to the standard literary language of Serbia, or the language which is recommended by the Committee for the Standardization of the Serbian language. “Croatian language” refers to the standard literary language of Croatia, or the language which is recommended by the Department of Croatian Standard Language. However, when comparing the languages in terms of identity, “Croatian” is the linguistic label that the majority of Croats identify with, while “Serbian” is the label that the majority of Serbs identify with. While the literary Serbian and Croatian languages are fixed standards, the Serbian and Croatian languages in terms of identity may encompass

many different dialects or conceptions. These terms will be used interchangeably for lack of a better alternative, but it is important to bear in mind the different interpretations.

It is also worth noting the distinction between “Serbs,” “Serbians,” “Croats,” and “Croatian.” Serbians and Croatians are citizens of the two countries, while Serbs and Croats are members of the ethnic groups associated with the country. Since the scope of this thesis is concerned with the effects of government policies on its citizens, and since it is more difficult to differentiate by ethnic group than nationality, this thesis will be comparing identity among “Serbians” and “Croats,” and word choice will reflect that. However, particularly in chapter three, there will be substantial comparisons of the ethnic group, and the phrases “Serb” and “Croat” will be used.

Chapter One: Literature Review

The ambiguous boundary between “dialect” and “language” points to the immense difficulty of categorizing languages. Even among the people considered “speakers” of a language, there are regional variations and differences in idiolects that render any distinctions arbitrary. In an influential 2007 essay, Makoni and Pennycook expand on this concept by arguing that all languages and conceptions of language are inventions of the state. The creation of language and traditions are part of the “constructed histories” of nation-states, used for “the creation of a past into which the present is inserted.”² Moreover, the notion of singularity and countability of languages that we today espouse is predicated on an ideology of territorialization which is inherently flawed due to its failure to recognize variation within and between states.³

By constructing this shared story, language becomes a circular tool for reinforcing nationalist ideas. In *Language, Ethnicity, and the State*, Camille O’Reilly argues that, because language has become an important marker of the borders between societies and states, the ideas of state, nation, and language become circular and self-evident: “each language group must constitute a nation, each nation should have a state, each state should have just one language.”⁴

² Makoni, Sinfree, and Alastair Pennycook. “Disinventing and Reconstituting Languages.” *Disinventing and Reconstituting Languages, Multilingual Matters*, 2018, p. 8, <https://doi.org/10.21832/9781853599255-003>.

³ *Ibid.*, pp. 10 - 11.

⁴ O’Reilly, Camille. *Language, Ethnicity, and the State*. Palgrave, vol. 1, 2001, pp. 8-9.

Although the idea of a “language” may be constructed by the state, it is still relied upon by the state’s citizens, who use linguistic identity as a “twin skin” to ethnic identity. Gloria Anzaldúa famously wrote that a language is something for people to “connect their identity to,” for “communicating the realities and values true to themselves.”⁵ Serbian linguist Ranko Bugarski also writes about the twin skin in regard to language policy in the former Yugoslav states, saying that “any notion of identity presupposes demarcation of the collective or personal self from the non-self, that is, the Other.” This delineation, in turn, “necessarily implies some kind of boundary” in language that people are drawn to in order to build their own identities.⁶

The phenomenon of both people and nations being incentivized to demarcate their language has far-reaching implications for national identity. Bugarski argues that language plays two roles: first, as a passive issue exploited by nationalists, and then as an active instrument for creating war propaganda and interethnic vitriol.⁷ Furthermore, the dominant group in a given nation-state is, through language, “accorded a special place, inevitably at the expense of minorities.”⁸ As David Laitin writes in his seminal work *Politics, Language, and Thought*, this means that some groups are put in a more advantageous position to

⁵ Anzaldúa, Gloria. “How to Tame a Wild Tongue.” *The Mexico Reader*, Duke University Press, 2022, p. 34, <https://doi.org/10.2307/j.ctv2rr3g8m.12>.

⁶ Bugarski, Ranko. “Language, Identity and Borders in the Former Serbo-Croatian Area.” *Journal of Multilingual and Multicultural Development*, vol. 33, no. 3, 2012, p. 229, <https://doi.org/10.1080/01434632.2012.663376>.

⁷ Bugarski, Ranko. “Language, Nationalism and War in Yugoslavia.” *International Journal of the Sociology of Language*, vol. 2001, no. 151, 2001, p. 70, <https://doi.org/10.1515/ijsl.2001.048>.

⁸ Bugarski, Ranko. “Language Policies in the Successor States of Former Yugoslavia.” *Journal of Language and Politics*, vol. 3, no. 2, 2004, p. 201, <https://doi.org/10.1075/jlp.3.2.04bug>.

take part in politics than others, allowing a “privileged group” to perpetuate language policy in their favor, and thus altering the arena further to their advantage.⁹

Following the dissolution of Yugoslavia, the self-perpetuation of language demarcation became particularly salient in Croatia. The American and Croatian linguists Keith Langston and Anita Peti-Stantić write that once the war was over, the Croatian language was considered to be in an internationally uncertain position, and therefore had to be defended from Serbian speakers living in and near Croatia.¹⁰ In Yugoslavia, the idea that Serbian and Croatian were just two dialects of the Serbo-Croatian language was a “key factor in promoting national cohesion,” but once Croatia became its own country, the assertion of Croatian as a language separate from Serbian was considered crucial for Croatia to be recognized as an independent nation with the right to self-determination. This idea was particularly based on the prevailing ethos of contemporary nationalism, which holds that “nation = language = territory = state,”¹¹ so it was a matter of survival for Croats.

To many Croats, Serbia’s insistence that Serbian and Croatian can be categorized as one language is an inherently imperialist idea tied to the hyper-nationalist idea of “greater Serbia” and stemming from illegitimate unification efforts in the nineteenth century. Meanwhile, to many Serbs, Croatia’s

⁹ Laitin, David D. “Politics, Language, and Thought: The Somali Experience.” *Politics, Language, and Thought: The Somali Experience*, 1977, p. 4.

¹⁰ Langston, K., and A. Peti-Stantić. *Language Planning and National Identity in Croatia*. 1st ed., Palgrave Macmillan UK : Imprint: Palgrave Macmillan, 2014, p. 6.

¹¹ *Ibid.*, p. 6.

insistence that Serbian and Croatian are separate languages is foolish, subversive separatism, reminiscent of the linguistic purism promoted by Croatia's fascist Ustaše government during World War II. This difference is realized in government policy: Serbia maintains a "neutral legal stance on variations" in language,¹² such as which script to use out of the Cyrillic and Latin alphabets, while Croatia upholds stricter standards, such as a "consistent requirement of the use of the Latin script."¹³

In *Language and Identity in the Balkans*, American linguist Robert Greenberg elaborates on the different ways language policy can be manipulated to further nationalistic goals by identifying three models through which ruling parties can use language to either unite or exclude populations.¹⁴ The "pluricentric unity" model is most akin to the Yugoslav example, wherein multiple cultural centers were allowed to coexist, each with their own literary and language norms, albeit in Yugoslavia this unity was "precarious at best" and "never truly embraced" by the competing ethnic groups.¹⁵ The "centrally monitored model" is similar to modern-day Croatia, with an academy, state-sponsored institute, or government ministry bearing responsibility for language planning, and the "government-imposed language unity" model entails direct totalitarian control over language policy.¹⁶ Similar to Langston and Peti-Stantić, Greenberg argues that Croatia's pivot to the centrally monitored model is out of a desire for

¹² Rice, Eric A. *Language Politics in Bosnia, Croatia, and Serbia*. 2010, p. 23.

¹³ *Ibid.*, p. 18.

¹⁴ Greenberg, Robert D. "Language and Identity in the Balkans." *Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration*, Oxford University Press, 2008, p. 19, <https://doi.org/10.1093/acprof:oso/9780199208753.001.0001>.

¹⁵ *Ibid.*, p. 18.

¹⁶ *Ibid.*, pp. 20-23.

“contrastive self-identification,” a behavior characterized by the desire of the Croats to underscore unique Croatian linguistic features in order to differentiate themselves from the Serbs.”¹⁷

Langston and Peti-Stantić note that this differentiation has taken the form of “traditional Croatian terms” being “reintroduced to replace forms that were employed during the Yugoslav period,” particularly in fields of government administration and the military.¹⁸ Instead of borrowing foreign words, Croatian linguists have traditionally preferred to *calque*, or create new words using the morphology of the Croatian language, and this attitude is often extended to the language itself. Any elements of Croatian that are considered extraneous, especially those which are associated with Serbian, are changed to a more Croatian form.¹⁹ The goals of this ongoing effort are to bring the language back to a “pure” state before it was imbued with unwanted Serbian influence, and to heighten the contrast between Serbian and Croatian norms so as to better support the idea that the languages are separate.²⁰ Croatian language policy in particular was notably unprecedented in the rapidity of its implementation and the level of effort the state committed to differentiate itself from Serbian.²¹

¹⁷ Ibid., p. 19.

¹⁸ Langston, K., and A. Peti-Stantić. *Language Planning and National Identity in Croatia*. 1st ed., Palgrave Macmillan UK : Imprint: Palgrave Macmillan, 2014, p. 13.

¹⁹ Ibid., 12.

²⁰ Ibid., 4.

²¹ Ibid., 13.

Bugarški notes that, “in contrast to its congeners, Serbian made no officially sponsored or approved moves to demonstrate its difference.”²² During the Yugoslav years, Serbia comfortably enjoyed the status of being the largest nation in the federation and could afford to be tolerant and broad-minded in linguistic matters because, as a rule, large nations have less reason to insist on their distinctiveness than their smaller neighbors, which may feel in danger of occupation or assimilation and therefore have greater motivation to separate their linguistic identity.²³ This is exactly what happened with Croatia, first with the Croatian Spring in the late 1960s and then following its split from Yugoslavia. Serbia had no need to change - in fact, it was actively disincentivized from it, because the similarity of Serbian and Croatian is a justification for unity. On the contrary, Croats had to move to differentiate their language as much and as quickly as possible from Serbian or Serbo-Croatian in order to protect their national identity. Ironically, this also means that, in its rush to join the West, Croatia purged its language of many Westernisms that would make that more possible, while Serbia, a nation that is more allied with Russia, retained the international terms that Croatian did away with.²⁴

In *Language, Nationalism, and the Yugoslav Successor States*, Robert Greenberg further explores the differences between Serbian and Croatian language policy, describing the “purist tendencies” of Croat linguists to cleanse

²² Bugarški, Ranko. “Language, Identity and Borders in the Former Serbo-Croatian Area.” *Journal of Multilingual and Multicultural Development*, vol. 33, no. 3, 2012, p. 231, <https://doi.org/10.1080/01434632.2012.663376>.

²³ Bugarški, Ranko. “Language, Nationalism and War in Yugoslavia.” *International Journal of the Sociology of Language*, vol. 2001, no. 151, 2001, p. 74, <https://doi.org/10.1515/ijsl.2001.048>.

²⁴ *Ibid.*, p. 84.

“their language of real and perceived Serbianisms,”²⁵ while “*status quo* linguists” in Serbia have continued the Yugoslav tradition of linguistics, and thereby successfully maintained the tolerant “language unity model” which was the official policy under Tito.²⁶ While Serbia never fully committed to the fate of Serbo-Croatian, Croat linguists endeavored to implement their “centrally monitored unity model” through substantial government interference and published numerous polemical works demonstrating that Croatian is indeed a different language from Serbian.²⁷

The uncertainty on Serbia’s behalf and consequent lack of “significant changes” is partially due to the country being “ill prepared for the new linguistic order”²⁸ and partially due to competing factions of linguists in Serbian academia. The “*status quo*” linguists believe that Serbian should evolve naturally, while “neo-Vukovite” linguists advocate for a return to the “pure principles of nineteenth-century Serbian language reformers” and “Orthodox” linguists fight for an extreme “Orthodox Serbian” version of the language.²⁹ So far, the *status quo* linguists have “maintained a position of dominance” in upholding the pluricentric unity model, but as nationalist politicians prevail in Serbia, there may

²⁵ Greenberg, Robert D. “Language, Nationalism and the Yugoslav Successor States.” *Language, Ethnicity and the State*, Palgrave Macmillan UK, p. 27, https://doi.org/10.1057/9781403914187_2.

²⁶ *Ibid.*, 36.

²⁷ *Ibid.*, 28.

²⁸ Greenberg, Robert D. “Language and Identity in the Balkans.” *Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration*, Oxford University Press, 2008, p. 86, <https://doi.org/10.1093/acprof:oso/9780199208753.001.0001>.

²⁹ *Ibid.*, 65.

be a shift to Orthodox linguistics.³⁰ This moment in time may be pivotal for further research on the divergence of Croatian from Serbo-Croatian because it precedes any serious divergence in Serbian - something that may not always be the case.

The evolution of the aforementioned polemical works is an incredibly useful resource for study of Croatian language policy. The foreword to a 1999 Croatian dictionary states that Serbo-Croatian “did not and does not exist as a concrete and spontaneous language... every text is realized either as a Croatian or Serbian text and every communication is realized either as a Croatian or Serbian communication. There is no text or communication that would simultaneously be both Croatian and Serbian.” It then elaborates that “Croatian standard language differs from Serbian in its entirety (as a system from a system) at all linguistic levels... they developed and took shape in different cultural and civilizational contexts” and are “independent linguistic expressions.”³¹ A 1997 language guide used as a reference in Croatian media and literature acknowledged that this mentality constitutes linguistic purity, but contended that this was not bad like conceptions of ethnic purity; rather, “striving for linguistic purity is a permanent characteristic of the Croatian language.”³²

³⁰ Greenberg, Robert D. “Language, Nationalism and the Yugoslav Successor States.” *Language, Ethnicity and the State*, Palgrave Macmillan UK, p. 36, https://doi.org/10.1057/9781403914187_2.

³¹ Barić, Eugenija., et al. *Hrvatski jezični savjetnik / c Eugenija Barić ... [et al.] ; u redakciji izvršnog uredništva Lane Hudečak, Milice Mihaljević i Luke Vukojevića*. Institut za hrvatski jezik i jezikoslovlje : Pergamena : Školske novine, 1999, p. 9.

³² Dulčić, Mihovil. *Govorimo hrvatski: jezični savjeti*. Hrvatski radio : Naprijed, 1997, p. 8.

In her influential paper *Nesting Orientalisms*, Serbian Slavicist Milica Bakić-Hayden argues that examples of “more Western” Balkan countries like Croatia striving for purity from “more Eastern” Balkan countries like Serbia is part of an overarching tendency to create a gradient of backwardness stretching from Western Europe to the Middle East. Croatia often portrays Serbia as being “wrong” or “behind” because it is closer to the Orient, which builds identity through the creation of an “other” to contrast with.³³ There are many other reference works, mostly from the 1990s when such discourse was most intense, similarly defending Croatian as being completely separate from Serbian and arguing for “purity” from Serbianisms. Both Croatian and Serbian linguists acknowledge the ongoing language purification of Croatian, and discussions of the facts are not necessarily in disagreement, but the frameworks on how they view the morality of that purification differ.

Textbooks are particularly revealing, as they have to be updated every few years with the latest vocabulary. Langston and Peti-Stantić analyze their impact by examining individual word and grammar differences between Serbian and Croatian textbooks, noting that it is through the educational system that the Croatian government has most explicitly and systematically attempted to enforce usage of the Croatian literary standard and introduce the Croatian language as a unifying symbol.³⁴ English political scientist Vanessa Pupavac elaborates on this

³³ Bakić-Hayden, Milica. “Nesting Orientalisms: The Case of Former Yugoslavia.” *Slavic Review*, vol. 54, no. 4, 1995, pp. 917–31. *JSTOR*, <https://doi.org/10.2307/2501399>. Accessed 8 Mar. 2023.

³⁴ Langston, K., and A. Peti-Stantić. *Language Planning and National Identity in Croatia*. 1st ed., Palgrave Macmillan UK : Imprint: Palgrave Macmillan, 2014, p. 115.

in her overview of ex-Yugoslav educational policy surrounding language, writing that an obvious motif in language disputes is what standard language is taught in schools and that “school textbooks embody national identity at a particular moment.”³⁵ It’s particularly important in Croatia, she argues, because “the establishment of Croatian as a separate language is explicitly linked to the Croatian state-building process.”³⁶

Langston and Peti-Stantić explain that the necessity of these materials and the arguments justifying them arose out of a need to demonstrate the continued existence of the Croatian people as a discrete national group throughout history. According to them, through the presentation of certain narratives, the state is able to mythicize the past in a way that further fosters nationalism.³⁷ This is a notable departure from Yugoslav language policy, where unique expressions of language were highly political and any manifestations of “Croatian linguistic identity were seen as dangerously nationalistic and separatist, and consequently as threats to the unity of the Yugoslav state.” In the past, having a unified Serbo-Croatian language was important for creating a Yugoslav identity and curbing nationalist tendencies.³⁸ Throughout most of their history, Croats have perceived themselves as being speakers of a minority language, whether under the rule of the Austrians or the Serbs, and independence was their chance to finally carve out an ethos of

³⁵ Pupavac, Vanessa. *Politics and Language Rights: A Case Study of Language Politics in Croatia*. Springer Nature, 2003, p. 3.

³⁶ *Ibid.*, 2.

³⁷ Langston, K., and A. Peti-Stantić. *Language Planning and National Identity in Croatia*. 1st ed., Palgrave Macmillan UK : Imprint: Palgrave Macmillan, 2014, p. 251.

³⁸ *Ibid.*, 8.

linguistic uniqueness and continuity.³⁹ Miro Kačić, former director of the Institute for the Croatian Language and Linguistics, notably said in a 1996 interview that “the Croatian language, along with the Croatian army, is the guarantee of Croatian existence and identity.”⁴⁰ Makoni and Pennycook would also surely agree that the insistence on a separate Croatian language is paramount for creating a “constructed history” that builds national and linguistic identity.

Many linguists have written about the Sapir-Whorf hypothesis, or the idea that language can influence thinking. A few, such as Komarzyca and Fras, theorize that there is also a link between language policy and thought. How a government regulates a language affects that language, which in turn affects people. It affects people more directly through creating norms of behavior and a certain culture around language.⁴¹ This political-linguistic relativity may extend to the Croatian and Serbian example, where the variation in government behavior toward language could affect how people view those languages and, consequently, construct their linguistic identities. In fact, that may be a partial aim of those policies.⁴² There already may have been a social impact from this difference, namely that a “culture of correction” has emerged in Croatia in which people correct others on improper Westernisms or Serbianisms in the language. Turk and Opašić note that many Croats exhibit a “critical attitude towards loanwords” and

³⁹ Ibid., 5.

⁴⁰ Ibid., 4.

⁴¹ Komarzyca, Daniel, and Janina Fras. “Language and Politics in India and China: A Cross-Cultural Comparative Study.” *Polish Political Science*, vol. 49, no. 4, 2020, p. 11, <https://doi.org/10.15804/ppsy2020402>.

⁴² Rice, Eric A. *Language Politics in Bosnia, Croatia, and Serbia*. 2010, p. 35.

resist their introduction by calling out others who use non-Croatian terms.⁴³ This may take place in online forums, in schools, in conversations, or even more passively through a general attitude.

Greenberg writes that language is perhaps one of the most important markers of national identity among Slavic countries because it plays a crucial part in differentiating between neighboring ethnic groups that may share the same heritage or religion.⁴⁴ Bugarski refers to Croatian language policy as an example of “reductive” nationalism, which constrains language as much as possible to build one cohesive group, separate from an *other*, while the lack of an isolating language policy in Serbia falls in the category of “expansive” linguistic nationalism, claiming that “all these ‘new-fangled’ languages were in any case nothing but political inventions, actually being Serbian under different names.”⁴⁵ To Bugarski, these different ways of interpreting language paved the way for the disintegration of Yugoslavia, because the dichotomy of who speaks what “came to be manipulated by all the warring parties in an attempt to retain or conquer territories, purify them ethnically and enclose them within linguistic walls.”⁴⁶

The Polish and Croatian authors Boduszyński and Pavlaković argue that the linguistic enclosure of Croatia was a necessary result of the “foundational legitimacy” narrative of the country, or the story justifying how the state came to

⁴³ Turk, Marija, and Maja Opašić “Linguistic Borrowing and Purism in the Croatian Language.” *Suvremena Lingvistika*, vol. 34, no. 65, 2008, p. 83.

⁴⁴ Greenberg, Robert D. “Language, Nationalism and the Yugoslav Successor States.” *Language, Ethnicity and the State*, Palgrave Macmillan UK, p. 17, https://doi.org/10.1057/9781403914187_2.

⁴⁵ Greenberg, Robert D. “Language and Identity in the Balkans.” *Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration*, Oxford University Press, 2008, p. 231, <https://doi.org/10.1093/acprof:oso/9780199208753.001.0001>.

⁴⁶ *Ibid.*, p. 227.

be. Since Croatian independence is relatively new, the need to build foundational legitimacy is more salient than in Serbia, and fostered the attitude of exclusion toward language and other cultural remnants of Yugoslavia.⁴⁷ This attitude has driven purification efforts that further reinforce the narrative.

Because purifying tendencies in an ethnically heterogeneous region can threaten diversity, linguistic minority rights are a prevalent concern in Serbia and Croatia. Following the dissolution of the common state and wartime migrations, Bugarski identifies “new minorities” that emerged in both countries that were not previously important in the Yugoslav pluricentric unity model.⁴⁸ Each government is concerned about members of its own ethnic group living within the borders of the other, and at the same time recognition of their language as official brings about the question of whether those speaking the minority language should be granted minority status.⁴⁹ Many in these minority groups are clamoring for expanded language rights in education and media, and are certainly using media differently than the majority groups in each country, something which is very under-studied.

While there is discourse on language policy in relation to the media, there is very little analysis of the Internet and social media. Langston and Peti-Stantić

⁴⁷ Boduszyński, Mięczysław P., and Vjeran Pavlaković. “Cultures of Victory and the Political Consequences of Foundational Legitimacy in Croatia and Kosovo.” *Journal of Contemporary History*, vol. 54, no. 4, 2019, p. 799. *JSTOR*, <https://www.jstor.org/stable/26787449>. Accessed 7 Mar. 2023.

⁴⁸ Bugarski, Ranko. “Language Policies in the Successor States of Former Yugoslavia.” *Journal of Language and Politics*, vol. 3, no. 2, 2004, p. 200, <https://doi.org/10.1075/jlp.3.2.04bug>.

⁴⁹ Langston, K., and A. Peti-Stantić. *Language Planning and National Identity in Croatia*. 1st ed., Palgrave Macmillan UK : Imprint: Palgrave Macmillan, 2014, p. 6.

mention that, in Yugoslavia, media outlets were directly under the state control and were expected to observe policies promoting Serbo-Croatian as a single language,⁵⁰ but that obviously changed following Croatian independence. Now, in Croatia, most of the print media is concentrated in two large corporations, which have given rise to “new concerns” about its objectivity, and the dominant HRT television network is controlled by the national news agency, with appointments “criticized as being based on political considerations rather than qualifications.”⁵¹ There is only one page devoted to the Internet, and it’s an analysis of blog postings in 2004-2005.

DellaVigna, et al. 2014 examines the impact of radio media on Serbian and Croatian nationalistic behavior. It found that “peaceful relations between neighboring ethnic and religious groups depend in part on the content of media programming” and that “nation-building efforts implicit in the nationalistic content” of media “can have important negative spillovers on the persistence of peace.”⁵² However, this study and others of its nature are limited to the effects of media on nationalism, and the only discussion of media in relation to linguistic identity are constrained to literature about linguistics, which may not be as comprehensive. Once again, the DellaVigna study and others predate the social media era.

Since the advent of social media is relatively recent and there is not much literature addressing it, examining the effects of language policy through the lens

⁵⁰ Ibid., p. 249.

⁵¹ Ibid., p. 250.

⁵² DellaVigna, Stefano, et al. “Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia.” *American Economic Journal. Applied Economics*, vol. 6, no. 3, 2014, p. 130, <https://doi.org/10.1257/app.6.3.103>.

of social media would therefore contribute something new to the field, and Serbia and Croatia are the perfect two countries for this research. Since “in Serbia we find Serbian basically unchanged, and in Croatia a Croatian fairly stabilised as distinct,”⁵³ this is a perfect opportunity to use Serbia as a control, demonstrating what would have happened if Yugoslavia remained intact.

Meanwhile, the differences observed in Croatian can be used to analyze the impacts of Croatia’s centrally monitored unity model. This is preferable to using one of the other ex-Yugoslav countries (Bosnia or Montenegro) because, while those countries are also implementing language changes to differentiate themselves from Serbia, their policies have been less extreme and less controversial than those in Croatia. However, much of this thesis may be generalizable to those countries.

Overall, the existing literature paints a very nuanced picture of the state of language identity and language policy in Serbia and Croatia. In chapters 2 and 3, I will be applying much of it as a framework to situate and contextualize my own analytical findings, nuancing the literature further through social media analysis and field interviews.

⁵³ Bugarski, Ranko. “Language, Identity and Borders in the Former Serbo-Croatian Area.” *Journal of Multilingual and Multicultural Development*, vol. 33, no. 3, 2012, p. 232, <https://doi.org/10.1080/01434632.2012.663376>.

Chapter Two: Analysis

This moment in time is perfect to compare linguistic identity in Serbia and Croatia because, while the Croatian language has measurably diverged from the Yugoslav standard in the past thirty years, Serbia still has not affected any significant change in its language policy, meaning that it can be used as a control to compare the effects of language policy. This window may close, however, as Orthodox and Neo-Vukovite linguists who advocate for a centrally monitored language model are increasingly winning the attention of the far-right ruling party, in contrast to the status quo linguists who have prevailed since the dissolution of Yugoslavia.⁵⁴ The social media landscape, of course, is constantly evolving as well, and this is a pertinent moment to conduct research in before it further changes.

Methods

To explain how language policy affects linguistic identity, I first need to understand the state of language identity in the region. In this chapter, I will analyze statements written by Serbs and Croats about their own language. By comparing sentiments of those statements between the countries, I will extrapolate that A) Croats are more likely to exhibit polarized feelings about their language, B) Croats are more likely to view the Croatian and Serbian languages as separate, and C) Croats are more likely to believe that Serbian is a language than

⁵⁴ Greenberg, Robert D. "Language and Identity in the Balkans." *Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration*, Oxford University Press, 2008, p. 65, <https://doi.org/10.1093/acprof:oso/9780199208753.001.0001>.

Serbs are to believe that Croatian is a language. I will be using the country as an independent variable and people's attitudes toward language as dependent variables from which I will be able to draw conclusions about identity.

To better understand the differences of how Croatian and Serbian speakers view what language they're speaking, I turned to social media. Twitter is the perfect snapshot into the last ten years because it's rife with easily accessible text data that can be used to analyze public opinion. The time interval in which tweets are available also lines up well with party control in the countries: the Croatian Democratic Union has been in power for most of the last two decades and in Serbia the right-wing Progressive Party (*Srpska Napredka Stranka*) has been in power since 2012. Neither country has significantly altered language policy in that time, so the variable of government influence is not something that needs to be considered when drawing a correlation between policy and identity. As a forum for casual conversation, the app is also a unique window into how people think about their own language in a casual context.

To pinpoint how Serbian and Croatian citizens identify linguistically (and then generalize that to Serbs and Croats), I chose to consider all instances where they refer to the language they speak. Since "Serbian" and "Croatian" are functionally dialects of what has historically been referred to as the "Serbo-Croatian" language, how one refers to the language can be very revealing about their identity. Additionally, since "Serbo-Croatian" is often associated with Yugoslav history, there has also been a modern movement to start referring to the language as "naš," or "our" language. This is, Bugarski writes, seen as a

“conveniently ambiguous phrase” that can be used to take a neutral middle route instead of implying that the language belongs to a particular nationality.⁵⁵ It’s also commonly used to “diminish the differences among” the languages and as a non-controversial way for people living in diaspora to talk about their language.⁵⁶ Comparing incidences of which people say “naš” across the two countries can also provide insights into which citizens have a greater interest in the idea of a united language.

Although the word people use to refer to their own language - or the *endonym* - can speak volumes about identity, it’s also best considered in the context of how they refer to what they consider another language - the *exonym*. By understanding how people classify in-groups and out-groups, we can understand how they think about who they are as a nation. In this case, the words “Serbian” and “Croatian” can be either endonyms or exonyms depending on the country they’re used in, while “Serbo-Croatian” and “naš” are rare examples of *ambonyms*, a word I suggest to describe names used to lump together in-groups and out-groups.

Using the names people have for the language they’re speaking as a way to compare language groups is especially useful in former Yugoslavia, where regional “dialects” can vary more within a language than the “languages” themselves. Instead of classifying people through the way they’re speaking, which

⁵⁵ Bugarski, Ranko. “Language, Identity and Borders in the Former Serbo-Croatian Area.” *Journal of Multilingual and Multicultural Development*, vol. 33, no. 3, 2012, p. 225, <https://doi.org/10.1080/01434632.2012.663376>.

⁵⁶ Jelić, Magdalena Petrović. “Serbian, Croatian, Bosnian, or Montenegrin: Our Language.” *Serbonika*, <https://serbonika.com/blog/serbian-language/serbian-and-croatian/>.

isn't helpful in this region, we can instead turn to something much more important - their identity, which comes through in their word for their language.

Filtering by location in Twitter and then searching for specific keywords creates an excellent corpus for analysis, but the limitations of Twitter must be acknowledged. Twitter has a reputation in the Balkans for leaning slightly more liberally, but the user base is considered similar in Croatia and Serbia. Although there aren't any available statistics on Twitter demographics in Serbia and Croatia, it probably has younger audiences and may not be representative of the makeup of the country as a whole. Again, though the likely similar demographics between Serbia and Croatia serve as a useful basis for comparison and it can still be considered a microcosm from which to extrapolate conclusions about the larger environment. Additionally, the ease of access and wealth of raw information make Twitter perfect for this thesis. Since Twitter can only be used to filter by location and not ethnic group, I will be assuming that findings for "Croatian users" are generalizable to "Croats" and findings for "Serbian users" are generalizable to "Serbs," although the groups may not overlap perfectly.

On October 28, 2022, I used a tool named Apify to scrape over 800 tweets with the keywords *srpski jezik* ("Serbian language"), *hrvatski jezik* ("Croatian language"), *naš jezik* ("our language"), *srpsko-hrvatski jezik* ("Serbo-Croatian language") and *hrvatsko-srpski jezik* (Croato-Serbian language"). After cleaning the data to make sure at least one of the full phrases was in each tweet and separating it into data sets of tweets with their geographic location set in Serbia

and tweets with their geographic location set in Croatia, there were 702 tweets remaining, with the following distribution:

Table 2.1: Number of tweets about each language name in Croatia and Serbia.

	"srpski jezik"	"naš jezik"	"srpsko-hrvatski jezik"	"hrvatski jezik"	TOTAL
Serbian users	406	74	18	38	536
Croatian users	15	20	1	130	166

The only instances of *hrvatsko-srpski jezik* were from Serbia, and since it's functionally used in the same contexts as *srpsko-hrvatski jezik*, I merged them into that dataset. Looking at the raw numbers of tweets is not very useful because there are, of course, a different number of Twitter users in Serbia and Croatia. Just to put the volume of tweets into perspective, therefore, I created another table (Fig. 2) normalizing for population. In 2022, there were 188,900 Croatian Twitter users⁵⁷ and 402,800 Serbian Twitter users,⁵⁸ so since there are 2.13 times more Serbian Twitter users we can control for the number of people by multiplying the number of Croatian tweets by 2.13.

⁵⁷ Kemp, Simon. "Digital 2022: Croatia." *DataReportal*, Global Digital Insights, 15 Feb. 2022, datareportal.com/reports/digital-2022-croatia.

⁵⁸ Kemp, Simon. "Digital 2022: Serbia." *DataReportal*, Global Digital Insights, 15 Feb. 2022, datareportal.com/reports/digital-2022-serbia.

Table 2.2: Number of tweets about each language name in Serbia and Croatia, normalized for population.

	"srpski jezik"	"naš jezik"	"srpsko-hrvatski jezik"	"hrvatski jezik"	TOTAL
Serbian users	406	74	18	38	536
Croatian users	32	43	2	277	354

This is surprising because, even after controlling for population on Twitter, Serbian users speak about language at a much higher volume than Croatian users (150% as much). This could be for a number of reasons, including Croats fearing the “culture of correction” prevalent in their country, more events meriting discussion of language in Serbia (at least thirteen Serbian tweets, for example, were about a controversy surrounding the crown prince’s ability to speak the language), or something else.

Sheer number of tweets is only so useful, however. Finding the proportion of discussion happening about each language name across Serbia and Croatia could be revealing about how each population is talking about language, so I divided the number of tweets about each language by the total number of tweets in their countries to generate the following table:

Table 2.3: Proportion of discussion of language names in Serbia and Croatia

	"srpski jezik"	"naš jezik"	"srpsko-hrvatski jezik"	"hrvatski jezik"
Serbian users	75.8%	13.8%	3.4%	7.1%
Croatian users	9.0%	12.1%	0.6%	78.3%

Overall, these numbers are very similar, with Serbs talking about the Serbian language in a similar proportion to Croats talking about the Croatian language. The same is true for vice versa and discourse on the phrase *naš jezik*. This similarity is good because now I could turn to sentiment analysis to find out how the speakers diverge.

For this, I turned to the GPT-3 text-davinci-002 language model,⁵⁹ which was at the time the most accurate AI text prediction software available for public use. Through a python program,⁶⁰ I first prompted it to complete a translation of each tweet, which repeatedly proved to be more accurate than Google Translate and often better than I could do as a heritage speaker. I then followed this with an explanation of the tweet and finally a sentiment score. The intermediary steps between the tweet and the sentiment score were included to train the model to be more accurate by giving it a logical process to reach the sentiment conclusion. Tests on individual tweets consistently were more accurate with this method.

The program classified each tweet with one of the following scores:

⁵⁹ Temperature was kept at the default setting of 0.7 because it seemed ideal to generate useful results without being too repetitive or too random.

⁶⁰ My friend Raphael Rouvinov-Kats '22 helped me write the code.

- A. “Extremely positive,” meaning it actively advocated for people to learn or use the language in question, or that it was separate from the others.
- B. “Positive,” meaning it somehow acknowledged the existence of the language without actively advocating for it. Note that this does not mean that the tweet necessarily conveyed a positive sentiment. The model was trained to assume that even a casual mention of a language in an unrelated context was an implicit acknowledgement of the existence of that language, and thus “positive” about that language’s existence.
- C. “Neutral/Unclear,” meaning the model is unable to discern a sentiment. For example, if a tweet is quoting another person who uses a word for a language, then it does not necessarily reveal anything about the author’s opinion about that language, making the sentiment unclear.
- D. “Negative,” meaning the tweet actively advocated against the language or the idea that one of the languages was separate from the others.

The model was trained through a prompt followed by numerous example tweets.⁶¹ I also manually went through to verify each result, because GPT-3 returned a “not found” error for about 8% of the tweets and accidentally parsed some positive statements as negative. While using natural language processing for text classification is at this point probably more accurate than traditional means of

⁶¹ See appendices A and B for how the model was trained.

sentiment analysis, the predictive nature of the AI needs to be kept in mind.

However, the model's output made it much easier to manually go through and classify everything.

Sentiment Differences by Language

The following table illustrates sentiment differences across Serbia and Croatia for discussion of the Croatian language:

Table 2.4: Sentiment differences for tweets containing the phrase "hrvatski jezik"

Sentiment	Tweets from Serbia		Tweets from Croatia	
	Tweets	Percentage	Tweets	Percentage
negative	10	26.3%	8	6.2%
neutral/unclear	8	21.0%	10	7.7%
positive	18	47.4%	83	63.8%
extremely positive	2	5.3%	29	22.3%

It is unsurprising but nevertheless interesting to note that Croats were far more likely to actively advocate for the Croatian language or passively acknowledge the existence of the language, while Serbs were less likely to view it as a language. While about a quarter of Serbian users talking about Croatian implied that it shouldn't be classified as a language, only 6% of Croatian users

felt that way (a figure that may reflect the ethnic Serbs living in Croatia), and while 47% of Serbian users somehow implied that Croatian is a language, fully 64% of Croatian users somehow implied that they speak a Croatian language. Meanwhile, only 5% of Serbian tweets actively advocated for Croatian as a language, in contrast with 22% of Croatian tweets.

The most revealing results from the GPT-3 analysis, however, are found when comparing these tweet sentiments with those about the Serbian language:

Table 2.5: Sentiment differences for tweets containing the phrase “srpski jezik”

Sentiment	Tweets from Serbia		Tweets from Croatia	
	Tweets	Percentage	Tweets	Percentage
negative	19	4.7%	2	13.3%
neutral/unclear	9	2.2%	1	6.7%
positive	330	81.3%	12	80.0%
extremely positive	48	11.8%	0	0.0%

Among Serbian users, there is much less dispute on whether Serbian exists as a language. 93% of the tweets from Serbia either passively acknowledged the existence of Serbian as a language or actively advocated for it, compared with only 53% of the tweets from Serbia about the Croatian language. At the same

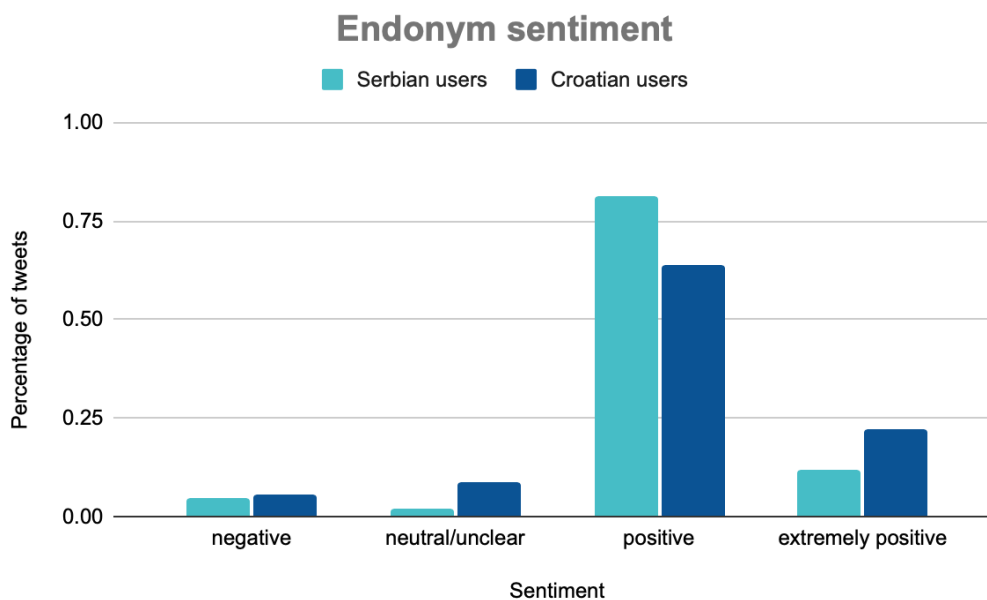
time, only 5% of Serbian tweets disagree with the idea of Serbian being a language, compared with 26% of Serbian tweets about the Croatian language.

Croats appear somewhat lukewarm toward the Serbian language. While 13% of Croatian tweets expressed negative attitudes toward its existence, compared to 5% of Serbian tweets about *srpski jezik* and 6% of Croatian tweets about *hrvatski jezik*, fully 80% of tweets acknowledged its existence - similar to the 86% of Croatian tweets either passively or actively acknowledging the existence of Croatian. This is probably because, according to the Croatian narrative, the existence of the Serbian language is a necessary requisite for the existence of the Croatian language. However, at the same time, no Croatian tweets actively argued that people should learn or use Serbian.

Endonyms and Exonyms

While it is useful to compare how Serbs and Croats have been talking about each language in isolation, it is a more practical insight into linguistic identity to compare how people of both nationalities are talking about their own language, or their *endonym*. For this, I compared instances of Serbs talking about the “Serbian language” and Croats talking about the “Croatian language”:

Fig. 2.1: Sentiment differences for endonyms in Serbia and Croatia



There appears to be a greater polarization between Croats writing about their own language than there is for Serbs. While 11% of tweets from Serbia spoke about the Serbian language in an extremely positive manner, 22% of Croatian tweets expressed extremely positive attitudes toward the Croatian language. Meanwhile, slightly more Croats tweeted negative things about the Croatian language than Serbs about the Serbian language. This could indicate that one group of Croatian speakers is experiencing a possible disconnect between what they speak and the language policies of their government, while another group of Croatian speakers readily embraces those policies. Serbs, not presented with a cohesive government stance on their language, are less likely to both think fervently and negatively about their linguistic identity.

This can be mathematically tested. Since extreme sentiments are being compared with moderate sentiments, it makes sense to test polarization by

combining the “negative” and “extremely positive” sentiments and comparing them to a combination of “neutral” and “positive” sentiments, which are not extreme.

Table 2.6: Highlighting extreme sentiments in endonyms

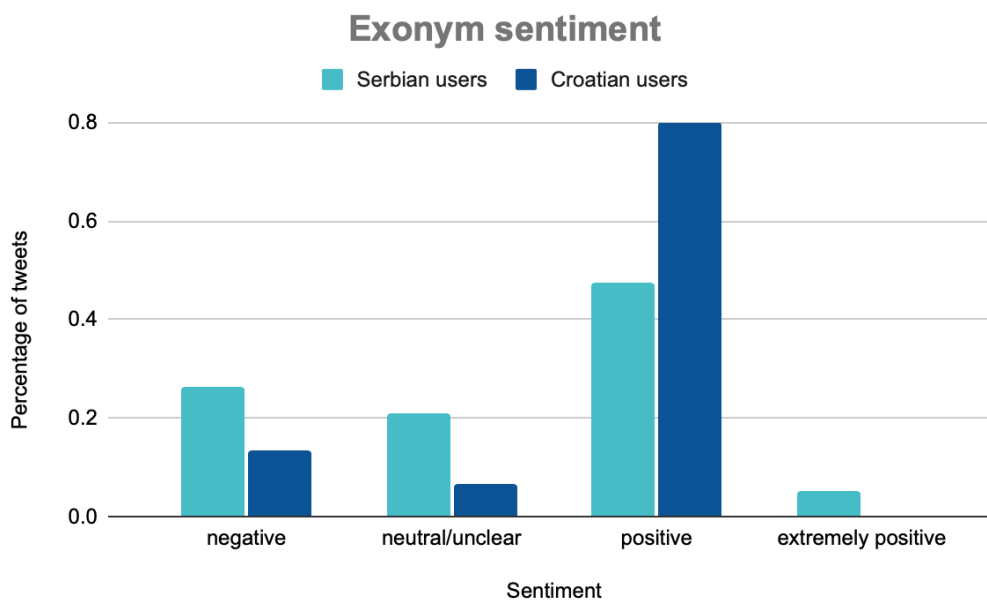
Sentiment	Serbian users	Croatian users
Neutral + Positive	339	93
Negative + Extremely Positive	67	37

This 2x2 frequency table is the simplest situation for statistical analysis. To test whether the proportion of extreme sentiments (negative or extremely positive) is higher in Croatia or Serbia, both Fisher’s exact test and a chi-squared test are appropriate. For very small samples, Fisher’s test is considered appropriate, while a chi-squared test works better when samples are very large. Since the counts in figure 7 fall somewhere in between, I did both tests.⁶² Both resulted in a p-value of 0.003, which means that the differences between samples are statistically significant at the commonly used 0.05 level.

We observe the opposite results when comparing exonyms, or how Serbian and Croatian speakers discuss the namesake language of the other country:

⁶² For the tests, I used an online statistics tutorial through Vassar College: <http://vassarstats.net/tab2x2.html>

Fig. 2.2: Sentiment differences for exonyms in Serbia and Croatia



Since people are more likely to talk about their own language, there is much less data for the exonym comparison than for the endonym comparison (53 tweets versus 536 tweets, respectively), so this should be carefully considered when looking at Figure 9. Nevertheless, the split is interesting.

While 13% of Croatian users made some kind of statement deriding or criticizing the existence of the Serbian language, twice as many Serbian users dismissed the idea of the Croatian language being independent. Meanwhile, 80% of Croatian tweets implicitly or explicitly acknowledged that Serbian is a language (indicated through the “positive” and “extremely positive” scores), while only 53% of Serbian tweets acknowledged that Croatian is a language.

This result is expected, because, again, the official stance of the Croatian government is that Serbian and Croatian are two separate languages existing concomitantly, while the Serbian government holds the opposite opinion that the

two are the same language, called “Serbian.” In the Serbian zeitgeist, many consider the Croatian language to be illegitimate, either because they feel that it broke off from Serbian or because they feel that the two languages are one.

One surprising result is that 5% of Serbian users wrote “extremely positive” things about the Croatian language, while no Croatian users wrote extremely positive things about the Serbian language. That 5%, however, is from only two tweets which look like they were probably written by Croats while their geographical location was set in Serbia (both tweets mentioned “Serbs” in a context that did not seem self-referential).

To test the hypothesis that the proportion of positive exonym sentiments (positive + extremely positive) is indeed higher in Croatia than in Serbia, I once again created a table from which I was able to do statistical tests:

Table 2.7: Highlighting positive sentiments in exonyms

Sentiment	Serbian users	Croatian users
Negative + Neutral	18	3
Positive + Extremely Positive	20	12

Because of the small sample sizes, only Fisher’s test applies. Under the one-sided hypothesis that Croatian users express more positive sentiments, the test result is 0.061, just above the customary 0.05 level, and the hypothesis is not accepted. On the other hand, if we exclude the two extremely positive sentiments

in the Serbian sample that were likely made by Croats, the count of 20 is reduced to 18, the test result is 0.045, and the hypothesis of differences is accepted.

These findings indicate only marginal or no statistical significance. However, small sample sizes lead to less precise statistical estimates. In view of the broader picture, I theorize that the observed differences are real and that Croats are indeed more accepting of the Serbian language as a concept than vice versa. Hopefully, this experiment can be repeated in the future with a larger sample size.

Attitudes Toward a Unified Language

In tweets using the phrase *naš jezik* (“our language”), it is once again Croatian users that exhibit a greater degree of polarization:

Table 2.8: Sentiment differences for tweets containing the phrase “naš jezik”

Sentiment	Tweets from Serbia		Tweets from Croatia	
	Tweets	Percentage	Tweets	Percentage
negative	10	13.3%	5	25.0%
neutral/unclear	6	8.0%	2	10.0%
positive	57	76.0%	11	55.0%
extremely positive	2	2.7%	2	10.0%

To collect these results, I slightly changed the prompt for the model. “Extremely positive” sentiment scores indicate that the author actively advocated for a unified language, “positive” scores indicate that they implicitly acknowledged the existence of *naš jezik*, while “negative” scores indicate that they somehow criticized or derided the idea of a unified language.

Based on the percentages, Croatian Twitter users were five times more likely to speak extremely positively and about twice as likely to speak negatively about the concept of the Croatian and Serbian languages being the same, possibly pointing to greater division within the Croatian population than the Serbia population.

To test this hypothesis, I again reduced the table to two grouped sentiments measuring extreme opinions:

Table 2.9: Highlighting extreme sentiments in ambonyms

Sentiment	Serbian users	Croatian users
Neutral + Positive	63	13
Negative + Extremely Positive	12	7

Due to the mid-sized sample, both Fisher’s Exact test and Chi-square test are again applicable in this situation. For the hypothesis that Croatia has a larger fraction of extreme sentiments than Serbia, Fisher’s test produced a value of $p=0.00006$. The Chi-square test, meanwhile, reported that the test value is

<0.0001. Thus, the observed differences are strongly significant, below both the customary 0.05 level and even the 0.01 significance level. This indicates that Croats are more polarized than Serbs about the idea of the Serbian and Croatian languages being the same.

Serbian Twitter users were also more likely to write positive things (either implicitly or explicitly) about the concept of *naš jezik* - 79% compared to the 65% of Croats using the phrase. In addition, even when adjusted for population, Serbia produced almost double the volume of tweets with the phrase than Croatia. This and the statistical test indicate that Serbs are more unity-minded about language than Croats, which would again fit in with the aforementioned national narratives.

This is perhaps also reflected in the greater proportional volume of tweets from Serbia using the phrase *srpsko-hrvatski jezik*, even though many of those tweets were about how people don't really use that phrase anymore, since it's more of an antiquated term associated with Yugoslavia. Of the 18 Serbian tweets in that small corpus, two expressed the extremely positive sentiment that people should use the Serbo-Croatian language, nine implicitly acknowledged the Serbo-Croatian language, and three were negative about the existence of the language, with four tweets containing neutral or unclear content. The one tweet from Croatia containing the phrase "srpsko-hrvatski jezik" was a negative statement translating to "Serbo-Croatian doesn't exist. There is the Croatian language and the Serbian language." There weren't enough of these tweets to really do an analysis for, but it's still interesting to note.

Overlap

After filtering through all the tweets, I discovered that 19 tweets from Serbia and 11 tweets from Croatia contained overlapping uses of the language descriptors. These were important to look at because, through explicit discussion and comparison of the names for the languages, they contain some of the most illuminating information on how people are thinking about languages in relation to each other. The first thing I did was manually sort them into whether they “implied division” between the Serbian and Croatian dialects (meaning they implicitly acknowledged or explicitly argued that the languages were separate), whether they were “pro-unity” (meaning that they implied or stated that the languages were the same), and whether the sentiment was neutral or unclear:

Table 2.10: Sentiment differences for tweets containing overlapping instances of language descriptors

Sentiment	SERBIA		CROATIA	
	Tweets	Percentage	Tweets	Percentage
implied division	6	31.6%	8	72.7%
neutral/unclear	7	36.8%	2	18.2%
pro-unity	6	31.6%	1	9.1%

Bearing in mind that this is an even smaller sample size than the exonym corpus, there is no point in applying statistical tests. However, we once again see that Serbian users appear more likely to use rhetoric that implies that the Serbian

and Croatian languages are the same, while Croatian users appear more likely to see the languages as different. This result is particularly notable due to it being from a context where the authors are comparing the languages and at least in a small way corroborates the earlier findings.

To better understand the circumstances under which people compare languages, I read through the thirty tweets and sorted them into the categories “education,” “media,” “government policy,” “social,” “history,” and “linguistic comparison,” or multiple of those categories if it applied. About the same percentage of tweets from Serbia (37%) and Croatia (36%) were simply comparing the linguistics of Serbian and Croatian, but Serbians were far more likely to be tweeting about the history of the language(s) or the government policy on language (47% and 16% of the tweets compared to 18% and 0% from Croatian users), which is surprising because these are traditionally more Croatian talking points. More Croatian tweets, meanwhile, complained about how their language was represented in the media (18% compared to 5%).

Conclusion

It would be exciting to see future analyses go further into depth into which topics Croatian and Serbian Twitter users discuss in addition to language, or how likely certain words are to show up in the same tweet as certain names for the language. If any other research uses natural language processing for this kind of sentiment analysis, especially if it involves larger data sets, it would be worth improving the code to prevent the “not found” error from returning. It would also

be a useful contribution to analyze Serbian and Croatian endonyms and exonyms in different contexts, such as Facebook, public forum sites, and surveys of the general population. This kind of analysis is generalizable to other linguistic situations with ambiguous names for the language, and the GPT-3 work is fantastic for any scenario requiring sentiment analysis in a foreign language.

Regarding this research question, the data suggests that:

- A) Croats are more likely to exhibit polarized feelings, both positively and negatively, about their own language than Serbs.
- B) Croats are more likely than Serbs to view the Croatian and Serbian languages as separate.
- C) Croats are more likely to believe that Serbian is a language than Serbs are to believe that Croatian is a language.

The next chapter of this thesis will corroborate these results through randomly sampled interviews and then attempt to explain the relationship between language policy and the observed effects in identity.

Chapter Three: Field Research

The sentiment analysis in the previous chapter provided a unique insight into the current condition of language identity in both Serbia and Croatia by demonstrating that Serbs are more unity-minded about language while Croats are more polarized about language. These differences were a very useful starting point to researching how language policy affects identity, because I was able to know which questions to ask that connect my findings to national structures.

This next portion of my thesis attempts to draw that link between language policy and language identity through field research - namely, key policymakers and scholars involved in Serbian and Croatian language policy, as well as interviews with randomly sampled citizens. I am trying to find the policy differences that are driving the differences in identity. I will be treating the political-linguistic situation in Serbia and Croatia as an independent variable while the differences in identity function as a dependent variable.

First, I will describe how the findings from my interviews corroborate my social media analysis that shows more polarized attitudes toward language in Croatia and more unifying attitudes toward language in Serbia. Then I will use the conversations with policymakers, contextualized through the street interview responses, to explain the main ways in which language politics differ between Serbia and Croatia, and how those differences affect identity. I will show that the structure of language institutes, the existence of national narratives, and the salience of identity all play a part in creating the observed effects of increased polarization in Croatia and increased unificatory attitudes in Serbia.

Methods and Street Interviews

Thanks to a grant associated with the Kenneth I. Juster Fellowship through the Weatherhead Center for International Affairs, I had the opportunity to travel to Zagreb and Belgrade, the capital cities of Croatia and Serbia, respectively, in early January 2023. I selected several different locations in the cities with the goal of obtaining responses from a wide range of locations. There, I stood in front of supermarkets and conducted open-ended surveys in the local language with random passers-by. I chose supermarkets because they were likely to experience steady foot traffic from a diverse range of people who live in the area. After recording age and where the respondents were from, my first three questions (discussed below) focused on confirming my social media findings. For my next two questions (“what influences how you think about your language?” and “how have politics and the dissolution of Yugoslavia affected your language?”), I let my respondents answer any way they wanted, with the goal of identifying any connections that existed between identity and language policy. I ultimately spoke with 28 Croats and 20 Serbs across all age categories.⁶³

The most striking observation I had was that the street interviews did in fact seem to corroborate my previous analytical findings. For example, when asked for the name of the language they speak, 15% of Serbs said *srpsko-hrvatski*, or “Serbo-Croatian,” compared to none of the Croats interviewed.

⁶³ See appendix C for sample interview guide.

Table 3.1: Responses to “what’s the name of the language you speak?”

Response	Zagreb	Belgrade
srpski	0	85%
srpsko-hrvatski	0	15%
hrvatski	100%	0

This mirrors the greater percentage of Serb Twitter users who were found to have positive sentiments toward the idea of Serbian and Croatian being a unified language. On the whole, these numbers were fairly similar to the proportion of discussion of each language name in Table 2.3, but some differences did exist. Nobody said *naš jezik* was the name of their language, which is unsurprising, because that’s mostly an informal shorthand way of talking about it. Meanwhile, it was surprising to learn that 15% of the Serb respondents used *srpsko-hrvatski*, because many consider it an antiquated term, but it’s possible that people are either more likely to use that phrase in person or that my question primed them to think about the kind of responses I might have been looking for. There did not appear to be significant variation across categories other than nationality: all of the *srpsko-hrvatski* responses were from different age groups, across different locations in Belgrade.

Likewise, when asked how they perceived the differences between Serbian and Croatian, Serbs were far more likely to categorize them as the same language:

Table 3.2: Responses to “how would you describe the difference between Serbian and Croatian?”

Response	Zagreb	Belgrade
They’re the same language	11%	80%
Similar, but with differences	68%	10%
They’re different languages	21%	10%

Much like the sentiment analysis finding how people talk about their language online, there once again appears to be more polarization among Croats on how they classify their language. In their responses, 68% of them notably acknowledged that their languages are similar, but stopped short of saying that they were the same, compared to 10% of Serbs. 21% classified them as different, compared to 10% of Serbs. Serbs were also more monolithic in their answers, with an overwhelming 80% of respondents claiming that Croatian and Serbian were the same language, compared to just 11% of Croats.

Finally, when asked about their thoughts on the use of the ambonym *naš jezik* (“our language”) as a unified term that can be applied to both languages, Croats were again more polarized as Serbs proved to view it more positively. Whereas half of Serbians were in favor of the term, only 18 percent of Croats said likewise; and while only 5 percent of Serbians were against the use of the phrase “our language,” more than 40 percent of Croats were.

Table 3.3: Responses to the question “what are your thoughts on the use of the phrase “naš jezik” for both Serbian and Croatian?”

Response	Zagreb	Belgrade
In favor	18%	50%
Neutral/Unsure	39%	45%
Opposed	43%	5%

Many Croats also seemed initially confused at the question, and it became apparent that the phrase was not widely used among locals, appearing primarily in expat communities. Most Serbs, however, were familiar with the term and a plurality reported that they either used it or supported its use. Much like in Table 2.8 in the analytical chapter, Croats appear to have more polarized opinions while a greater percentage of Serbs had positive or moderate opinions.

The street interviews corroborated my findings in the analytical chapter: Serbs are more likely to support the idea of a unified *naš* or *srpsko-hrvatski* language, and Serbs are more likely to see Serbian and Croatian as the same language while Croats are more likely to see them as different. As with the first survey question, there did not appear to be significant differences in respondents across age or region of the city, although younger people seemed slightly less likely to say that the dissolution of Yugoslavia affected how they think about language. After identifying this, the next step was to find what political factors shape the observed differences.

In addition to the street interviews, I arranged to speak with several influential linguists and language policymakers in both countries - the first time such interviews have been incorporated into research on language identity in the region. These included the following:

- A) Milica Mihaljević, Director of the Department of Croatian Standard Language
- B) Željko Jozić, Director of the Institute for Croatian Language and Linguistics
- C) Sreto Tanasić, President of the Committee for the Standardization of the Serbian Language
- D) Milorad Dešić, retired member of the Committee for the Standardization of the Serbian Language
- E) Prvoslav Radić, professor at the Serbian Academy of Sciences and Arts who has written extensively on language policy in Serbia and Croatia.

I was also able to conduct brief email conversations with Professors Ranko Bugarski and Noam Chomsky, who have written extensively on the topic and were both signatories of the *Declaration on the Common Language*. Bugarski noted that my research results and the differences I ultimately noted between Serbian and Croatian linguistic identity “fully correspond to [his] own impressions.”

In addition to the survey respondents’ comments on the interaction between government and language, these interviews critically helped connect language identity with language policy. Through the discussions, I identified three

political mechanisms in Serbia and Croatia that affect linguistic identity: the structure of the language institutes, the national narratives in place, and the salience of identity.

Structure of Language Institutes

Serbian and Croatian language policy is similar in that both rely on the recommendations of centralized language institutes in their respective countries. Although there are some laws in place regarding nomenclature in government or educational settings, neither country has fully standardized regulations forcing the media or people to speak or express themselves in specific ways. Instead, these institutes provide suggestions where they think it prudent.

Beyond that, the structure of the institutes in the two countries is quite different. Serbian language policy is issued by the Committee for the Standardization of the Serbian Language, which operates under the auspices of SANU, the Serbian Academy of Sciences and Arts. Meanwhile, decisions about the Croatian language are made by the Department of Croatian Standard Language in the Institute for Croatian Language and Linguistics. Under Yugoslavia, both institutes existed to study “Serbo-Croatian” linguistics, and the Serbian institute largely remains as an extension of that objective, but in 1990 the Croatian institute was remodeled to study the Croatian language exclusively. This institute is smaller in size and independent of any larger academic body, but has a closer relationship with the government and media. I spoke to the directors of both institutes at length, and their reflections suggested that these differences

greatly affect the institutes' function and impact—which in turn shapes how language shapes identity.

Possibly as a result of its smaller size and less bureaucratic structure, the Institute for Croatian Language and Linguistics appears to accomplish far more of its objectives. According to Director Jozić, their Serbian counterpart is “slower in making decisions” because they are more conservative about the idea of changing language. Jozić believes that “a rational, realistic, balanced, modern approach to language” and a willingness to make decisive judgments leads to more people accepting the language and greater success with changing it than the conservative approach.

The Croatian Institute's greater ability to quickly and liberally affect language planning partially results from its independence. Jozić repeatedly brought this up, noting that it was incredibly important to him as a director. Unlike the Serbian institute, Croatian language researchers do not have to seek permission or funding to pursue research projects or make decisions. Director Tanasić of the Serbian committee acknowledged this difference, although he noted that some of the Serbian committee's projects are not financed by the parent Academy of Sciences. Another divergence in the power structure of the institutes is that, in Serbia, all media decisions have to go through Tanasić, who gives the final opinion on the issue. Jozić argues that the smaller size, looser hierarchy, and greater flexibility in the Croatian institute allows them to respond to the media more quickly and decisively.

Indeed, even the physical appearances of the institutes hint at their structural differences. The Croatian institute has its own building, which stands majestically on a row of ministerial edifices that all have Croatian flags flying out front. The Serbian institute is relegated to scattered offices, all of them connected to noisy college hallways that echo with the chatter of undergraduate students.

Age and technology also play a role in the ability of the institutes to affect change. The Serbian committee staff members are on average much older, as many of them are left over from the Yugoslav days. Meanwhile, many of the linguists currently working for the Croatian institute were hired in the 1990s when it transitioned into focusing on the Croatian language exclusively and adopted new projects like creating a standard Croatian dictionary. Jozić placed a lot of emphasis on this generational divide, repeatedly emphasizing his youth as a director and arguing that there exists a “generational problem” of older linguists wanting to do things the old way and not being open to suggestions, which again ties in to the conservativeness of the Serbian institute.

This generational problem manifests in part through utilization of the Internet in language planning. Both institutes have websites providing recommendations on style and word choice, but Jozić placed more emphasis than Tanasić on the importance of using digital media and the Internet to affect change. A chief objective of the Croatian institute is currently to make language resources and recommendations available to all Croats through cyberspace, and Jozić strongly believes that the Croatian language has already “successfully gone through this transition,” putting it on the right path for developing in an online

era. Meanwhile, he jokes, maybe his Serbian colleagues “don’t even know how to turn on the Internet.” It’s even possible that Serbia just recently missed a huge window for online language policymaking. Jozić talked about how during the pandemic use of the Croatian institute's online tools skyrocketed and expressed doubt that the Serbians were able to build up their online presence as successfully.

One of the ways in which the Croatian institute affects more change than the Serbian institute is that it exhibits greater policy aggressiveness by issuing more decisions on how individual words should be used. Tanasić noted that, despite the larger size of the Serbian committee, “we don’t give advice for every word, we have a general attitude.” The Serbian institute gives general, sweeping recommendations for how the Serbian language should be classified and treated, answering questions as they come up. The Serbian ethos, Professor Dešić says, is to “not change language too much because that would be dishonest.” This ideology is consistent with the “status quo” linguistics described by Robert Greenberg.

Linguists in the Croatian institute, however, believe that the only way to provide ethical stewardship of the Croatian language is by vigilantly protecting it from foreign loanwords through suggestion of calques (alternative words that translate the term directly into Croatian). The replacement of loanwords with calques also includes re-evaluating the usage of existing words of Serbian, German, English, and Turkish origin if there is a Croatian alternative. Director Mihaljević described this philosophy as the idea of things being “better in Croatian,” also the name of a website run by the Department of Croatian Standard

Language. This process, Mihaljević said, takes the form of finding a Croatian word “that is used but maybe not as popular” and proffering that instead of foreign alternatives when creating style guides and making recommendations to the media. Ultimately, the introduction of “more Croatian” words is meant to nurture a distinct Croatian identity.

A good example of how Croatian policymaking can impact individual words is the translation for the number “one thousand.” Two of the survey respondents in Zagreb independently brought up the fact that they used to say the “more Serbian” word, *hiljadu*, but with the new government they heard the “more Croatian” word, *tisuću*, used more and more in signage and media until they started using it themselves. “They bombard you to get into your head *tisuću*, so now even I say *tisuću*,” one of them explained. When I asked Mihaljević why *tisuću* suddenly seemed to gain prominence as a word, she said the Croatian institute has been advising journalists and official documentation to use it because it was “better in Croatian.” This policy, Dešić notes, has pushed some speakers of the northern Kajkavian dialect to start saying *tisuću* even though their Croatian dialect did accept and use *hiljadu* before. Making decisions about the national language standard may have unintentionally impacted the dialects within Croatia that individually differ more from the norm than even the Serbian standard does.

As noted in my literature review chapter, the Croatian insistence on calquing and using traditional Slavic roots has resulted in the curiously paradoxical result that, despite being more culturally aligned with the West, their language has evolved to be closer to a Slavic ideal. Linguists in both countries are

aware of the irony. Professor Radić jokes that, because Croatians “make identity in the form of archaisms” they are better guardians of Vuk Karadžić, the famous Serbian language reformer. He believes that the reason the Serbian government has not tried to similarly change language is that there is “an important difference between a nation that takes care of its identity because they know how they created it and a nation that believes they don’t have to worry about their identity.” To him, Serbia falls in the latter category, a concept that will be discussed in depth later.

Part of the reason the Croatian institute is able to implement changes like increasing the use of *tisuću* is that they have a better relationship with the media than the Serbian institute. Director Tanasić in Serbia complained that “newspapers that do not have state money” still write in the Latin script and generally don’t heed their suggestions. Tanasić also enumerated several circumstances of television and newspaper media failing to inform the institute about language style changes they made. Radić agreed, noting that media influence was stronger in Croatia than in Serbia because at “the language institute there they teach the media to use the media.” Indeed, in the examples of media interaction that Jozić and Mihaljević mentioned, they described their relationship with the Croatian media as the media frequently calling *them* about advice on which word to use. Even though neither institute regulates the media, there is a greater culture of institutional suggestions being heeded by the media in Croatia.

Both institutes play a role in word choice for educational textbooks and materials, but the Croatian institute updates them more frequently as they

continue to nudge their language in a more traditionally Croatian direction. Jozić noted that “it all starts in the educational system” and that one of his chief objectives was to “have an influence on what language will be used in the educational system.” Professor Dešić in Serbia criticized this attitude, arguing that Croatian “textbooks are made artificial by geography” because they falsely equate the idea that the borders of Croatia are coterminous with the Croatian language.

Finally, the relationship between the language institutes and their national governments differs. In the Serbian case, it can be strained at times. Tanasić further complained that “sometimes the state listens, sometimes it doesn’t” and that the Croatian institute has had stronger support from their state than the Serbian institute has thus far, which may further impact their respective abilities to affect change. Both appear to operate separately from the political party in power, although the government does choose the director of the Croatian institute. I could not find out how the Serbian director is chosen, but since it’s part of a university, it’s probably through a different method.

The differences in policy aggressiveness, ethos, and effectiveness of the institutes are inextricably linked with the construction of language identity. By making more changes to standard Croatian that are then picked up by the government and media, the Croatian institute is affecting the vernacular of everyday people, which then affects how they think about language. The fast pace of language change resulting from the structure of the Croatian institute also critically explains why there is greater polarization surrounding language in Croatia: some people feel disenfranchised, while others are able to feel more

patriotic. At the same time, the Serbian institute's "status quo" attitude of preserving the Yugoslav norm has helped serve as a justification for the increased sense of language unity in Serbia. Language policy in both countries is treated with the intention of affecting identity, and accomplishes just that.

National Narratives

The relative receptivity of Croatia's media and government to the suggestions of its language institute may be due in part to the differences in how people see the status of their own country. The importance of Croatia leaving the Yugoslav federation while Serbia continued to exist as a successor state cannot be overstated. All my interviews with the policymakers and random citizens pointed to a fundamental contrast in the construction and perception of national narratives surrounding identity and language.

Makoni and Pennycook argued that national language traditions are based on "constructed histories" tying the idea of language to a territory, and this appears true in different ways for Serbia and Croatia. Both sides acknowledge that Croatia views itself as fundamentally separate, while Serbia still believes that the countries share a language and identity. This was reflected in the survey responses. Director Tanasić phrased it particularly well: "from our side it's Serbo-Croatian, from theirs it's Croatian." Due in part to lingering Yugoslav ideas about language unity and in part to an expansionist desire to reunify in some way, Serb linguists have continued to make linguistic decisions on the premise that

they still need to influence the development of the Serbian language in the other ex-Yugoslav nations.

This means that they are less incentivized to make broad language changes, which would leave behind Serbian minorities in those countries. Director Jozić claims that Serbia's language policies are ultimately rooted in a continuing "need to proclaim the unity of Yugoslavia, a common state, to the detriment of its own Serbian identity." Instead of differentiating a national language identity as Croatia has done, which would be unnecessary, Serbia has focused on protecting and preserving Serbian abroad through Greenberg's pluricentric unity model. According to Tanasić, the Serbian institute prioritizes "protection of the status of the Serbian language in the entire Slavic world," and Radić also claimed that the attitude of the Serbian institute is to "fight for the preservation of the cultural heritage of the Serbian language" everywhere.

This is a topic of great concerns for Serb linguists and policymakers. Radić worries that there is an ongoing attempt to replace the Serbian linguistic space with another name, and Professor Dešić stresses that "it is our job to choose how to protect the Serbian language, that it does not become something else, that it does not disappear." Just as linguists in Croatia steadfastly maintain that Croatian is a different language from Serbian, Serbian linguists maintain that they are the same because they don't want to abandon Serbs living in Croatia.

This narrative of caring about the Serbian language in other countries likely explains why Serbs are more likely to use *naš* and respond that the languages are the same. This could be observed in the street interviews. One

Serbian respondent said that the Croatian government “tried to separate us by force and by language,” but have still not succeeded, and another defended his use of ambonyms by saying that “if we are not in the same country, we like to recognize each other.”

A very different story has played out in Croatia, where one respondent called the idea of *naš jezik* a “political lie.” In Jozić’s opinion, the biggest difference between Serbian and Croatian language policy is that “Croatians have always had the culture of the Croatian language and the culture of Croatian identity, and in Serbia they have the identity of the common space of Yugoslavia.” The Croatian linguistic community has adopted an attitude of isolationism and individualism with their language, choosing to focus only on the Croatian language as it is spoken in Croatia. Boduszyński and Pavlaković claim that this is to reinforce “foundational legitimacy” and build identity through a shared narrative of uniqueness and difference from an “other.” They elaborate by highlighting numerous non-language examples of how foundational legitimacy underlies the cleavages in Serbian and Croatian ideologies: Croatia has adopted exclusive economic and political policies, and educational resources reflect exclusive historical narratives.⁶⁴

While Serbian language policymakers openly admit to making linguistic decisions based on how Serbian is being used in other countries, Jozić holds that they only want to consider the Croatian language in isolation, similar to

⁶⁴ Boduszyński, Mieczysław P., and Vjeran Pavlaković. “Cultures of Victory and the Political Consequences of Foundational Legitimacy in Croatia and Kosovo.” *Journal of Contemporary History*, vol. 54, no. 4, 2019, p. 810. *JSTOR*, <https://www.jstor.org/stable/26787449>. Accessed 7 Mar. 2023.

Greenberg's description of a centrally monitored language model. Director Mihaljević agrees, saying that her committee tries to keep standard Croatian separate from Serbian because they choose to just "look at Croatian as a system" and "don't compare it to Serbian." While early Croatian textbooks from the 90s did compare it to Serbian, educational materials and style guides today only present the Croatian language by itself.

The attitude of Croatian individualism has been around for a while. One survey respondent in Zagreb emphasized that "even before the breakup of Yugoslavia I was defined as a Croat," and though they used the Serbo-Croatian language in the army, he still considered himself a Croat who had always spoken Croatian. Jozić echoed this attitude, claiming that the Croatian people are sensitive to and want to protect their language because it's an important part of their identity that they've been holding on to. He pointed out that more than 50% of Croatian words are of foreign origin and now there is a zeitgeist of protecting the language from external corruption.

As an example again, this explains why the Croatian language institute has pushed for use of *tisuću* instead of *hiljadu*. Replacing a word that is perceived as foreign with a word that is perceived as Croatian helps foster this sense of Croatian individualism. Meanwhile, in Serbia, language expansionists like Professor Dešić complain that such a decision leaves behind Croatian "speakers who still say *hiljadu*."

Croatian language isolationism also explains why Croats are both more likely to identify Serbian as a language than the other way around. "Nobody in

Croatia has a problem with the existence of the Serbian language, we all think it exists,” Mihaljević told me, in stark contrast to the Serbians who oppose the idea of Croatian being its own tongue. Simply put, since Croats have a tradition of viewing their language as discrete, they’re more likely to also see the Serbian language as discrete too. This makes up the difference between Bugarski’s reductive and expansive linguistic nationalism.

Professor Bugarski explained the differences in identity-related popular reactions surrounding language policy as “ultimately reducible to matters of social psychology” surrounding a narrative of a “Big Brother relation” between Serbia and Croatia. In his example, most Serbs see Croatia as a rebellious little brother, while Croats see Serbia as an oppressive older brother. Their policies, and the identity differences that result from them, follow from the narrative that they are in a precarious position that must be defended.

In addition to internal constructs of national narrative, external ideas about the nature of Serbian and Croatian may have affected the development of identity surrounding them. Professor Radić argues that even the label “Serbian language” is ultimately a Western imposition meant to ascribe Serbia’s language standard unto only the Serbian people. The exonym “Croatian,” meanwhile, is designed so that it can only be interpreted as applying to Croatian people. Radić says the word “Serbian” as it is used now “only applies to half of the Serbian language,” because it was constructed to not include the language of other ethnic groups. Just the name of the language is important to how people are predisposed to linguistically identify, because ethnic or national identity becomes tied up with the

“twin skin” of linguistic identity and the prevailing narrative ethos behind foundational legitimacy is “nation = language = territory = state.”

Due to a long history of imperial oppression, Serbia has always had a very wary attitude toward the West, but this has particularly been true since the 1999 NATO bombings. Radić claims that the break-up of the Serbo-Croatian language occurred along with the break-up of Yugoslavia because of European and American interests meddling by creating artificial ideas about new languages in the region. Foreign influence is felt by some members of the public, too: one of the Belgrade respondents said that the main factor that affects how he uses language is how it is perceived by foreigners. Not wanting to sound “aggressive” to outsiders, he both chooses his words and talks about his language in a non-offensive way.

National narratives are propagated by government actions, whether through small-scale policy changes, like standardizing *tisuću*, through the way things are presented in the media, or through broad government stances, like whether Serbian and Croatian are dialects or languages. These narratives then feed into identity by giving groups a way to either differentiate themselves from an “other” or unite themselves around a shared story.

Salience of Identity

The increased polarization of opinion on language status in Croatia may also be due in part to identity simply being more important in Croatia. Croatia has been independent for less time than Serbia, which had a prior history of being

independent and is currently seen as an extension of Yugoslavia. Since Croatian national identity is newer and not as established as Serbian identity, there is more of a need for Croatia to differentiate itself from the Yugoslav language standard through what Greenberg labels “contrastive self-identification.” Policymakers in both countries acknowledge this difference. Professor Jozić noted that, for Croats, “nationhood is a new category,” which means they need new ways of looking at identity and what makes everybody members of a certain community. For him, it was “language which clearly determined it.” Professor Tanasić also speculated that “language is maybe more important for identity there in Croatia” both because of the recency of its development and because Croatia is more in danger of experiencing erasure through pan-South Slavic ideas. In his brief email to me, Professor Chomsky, too, theorized that Croatia’s language policies may “reflect deeper commitments to national identity.”

There is a longstanding Croatian tradition of perceiving existential threat to language which very clearly has impacted attitudes toward language and language policy. Jozić repeatedly brought up that the Croatian language has experienced linguistic repression through the entirety of its existence. Before Yugoslavia, it was constantly suppressed under the Italians, the Austro-Hungarians, and the Ottomans. People speaking Croatian couldn’t be hired in some circumstances, and the language wasn’t properly represented in education, media, and other forms of recognition. Yugoslavia was a continuation of the same story. Multiple Croat respondents highlighted that despite the branding of their language being a “Serbo-Croatian,” Serbian terms were imposed

in public service, particularly through military nomenclature. The threat that “someone wanted to take [their language] away” is how the “most important part of identity was born” in Croatia. Mihaljević concurred, pointing out that, while most Croats in Yugoslavia retained a Croatian identity and the idea that they were speaking Croatian, the first time they could express that identity freely was over the last thirty years. One Croatian survey respondent noted that the newness of their country has “influenced the awareness of the Croatian nation as our own nation” as well as “the need and necessity to maintain what is ours.”

The fact that the Croatian language is changing faster than the Serbian language is both a byproduct of Croatia’s greater emphasis on identity and a further catalyst of identity importance. When I asked Tanasić to explain why Serbs were less polarized about their language identity, he said that the Serbian language relative to Croatian is “still dormant, and that’s why we have such great support in the people.” Croatia, on the other hand, is making a lot of changes very quickly, which may disenfranchise some people or leave them behind.

Serbia’s reticence to make fast changes to language is to some degree a leftover artifact of the Yugoslav attitude of not discussing language in their country. Under the ideas of “brotherhood and unity,” people were strongly discouraged from identifying any factors of division between them. This initially made it difficult for the Serbian institute to do its job. According to Tanasić, it was a tradition in Belgrade that you “couldn’t talk a lot about your language in the 90s.” Radić agrees that “Serbs lost their identity after Tito’s communism was introduced and supported.” To this day, Tito’s legacy seems to have contributed to

the lessened emphasis on identity in Serbia, while Croatia eagerly started distancing themselves from Yugoslav ideals as early as they could.

The greater sense of identity in Croatia has manifested through the increased reliance on calques and the “more Croatian” terms like *tisuću* proffered by the Croatian institute. Durability and longevity are perceived as increasing linguistic identity, and Tanasić claims that Croatian policymakers are trying to convey that longevity and historical continuity of their language through the reintroduction of historical, “Croatian” words. It is the Croatian tradition to “use a lot of calques” because they have historically been afraid of “Germanization, Magyarization, or Serbianization.” Radić says that these policies are used to assimilate or expel ethnic Serbs in Croatia by closing out the Serbian framework and language that was previously available to them, as well as creating noticeable differences for Croats through linguistic engineering. Clearly, the calquing is important for identity.

Croatia’s greater reliance on calques and relative policy aggressiveness reinforce the importance of identity in the country. By implicitly or explicitly being taught that the way they speak is important to their identity, many Croats internalized that through generating exclusive attitudes toward language. The disconnect between those who did and those who didn’t is perhaps one reason why Croats are more polarized about language.

Retrospective

Serbia and Croatia occupy two very different moments in their cultural, social, and linguistic histories which are in turn driving the directions of their language policies and the structure of their language institutes. In this chapter, I show that the primary mechanisms through which language policy affects language identity in Serbia and Croatia are 1) through the different structures of the Serbian and Croatian language institutes, 2) through the construction and propagation of different national narratives, and 3) through a greater emphasis on identity in Croatia than in Serbia.

Despite this, it was consistently surprising how cyclical everything felt. As Camille O'Reilly writes, "language becomes a circular tool" and the aforementioned differences in national narratives, salience of identity, and structure of institutes are all inextricably related to each other. The institutes exist the way they do because of the stories in place, which exist because of the varying strength of identity. At the same time, the strength of identity exists because of the stories in place, which are driven by decisions made at those institutes. Both countries are locked into a self-perpetuating loop of politics and identity which affects everyday people in ways they often don't consider.

Regardless of the opinions they actually held about their language and the political touchiness surrounding language policy, officials and citizens in both countries consistently agreed about which differences existed and how the other group perceived them. The only actual points of contradiction were normative, but when asked to objectively describe the situation; everyone responded with a

matter-of-fact “this is how they feel, this is how we feel, and why.” In Zagreb, Croatian and Serbian were considered separate languages, and in Belgrade, they were considered similar dialects. This difference in identity is ultimately a product of the other differences found in language policy.

Conclusion

As the Croatian language standard grows further apart from Serbian, the question of whether “Croatian” and “Serbian” are the same language will become even more important and divisive in the region. Proponents of language unity will grow more frustrated that a separate Croatian language identity is stabilizing, while Croat nationalists will have more evidence to point to when arguing in favor of their linguistic uniqueness.

In Chapter Two of my thesis, I identified that A) Croats are more likely to exhibit polarized feelings about their language, B) Croats are more likely to view the Croatian and Serbian languages as separate, and C) Croats are more likely to believe that Serbian is a language than Serbs are to believe that Croatian is a language.

In Chapter Three, I sought to explain how the observed differences in Serbian and Croatian identity are driven by language policy, and I found that the political factors affecting the identity can be explained through the mechanisms of A) the structure of the language institutes, B) different national narratives, and C) differences in the salience of identity.

These results corroborate the literature review and are therefore expected, but far more nuanced than anticipated. The cyclical nature of linguistic nationalism was particularly striking. As time goes on, the cycle of identity driving policy and policy driving identity will surely continue. The Croatian language institute will continue to make decisions on how individual words are brought into the standard language, while the Serbian language will continue to be

inundated with an influx of English loanwords. The story of Croatia being separate will be reinforced with time and an increasingly separate language, while the Serbian institute will desperately continue trying to influence how their language is used by ethnic Serbs beyond their borders. The idea of language lending Croats their identity will grow stronger as it emerges as a more distinct way for them to differentiate themselves.

In a time when this topic is becoming more and more important, this thesis contributes an original perspective by uniting and synthesizing discussion of Serbian and Croatian language politics with language identity in a new way. While there have been plenty of individual writings on how national narratives and salience of identity affect language politics in the region, this thesis is the first to isolate those factors in this way.

Moreover, there has been little to no discussion of how the structure of the language institutes affects language policy and, in turn, language identity. This thesis is the first to elaborate on that through incorporating interviews with key players in Serbian and Croatian language policy. Both the professional and personal interviews add a new method of research to existing discourse, which has largely been from theoretical perspectives that have not utilized field research.

This thesis is also the first to treat the word people use for their language as an indicator to conduct analyses on identity, as well as the first to incorporate social media sentiment analysis and natural language processing into discussion of language identity in Serbia and Croatia. The unique mixture of quantitative and qualitative discussion provides a novel lens that will be helpful to future research.

One potential limitation of this thesis is the amount of data available for analysis. The small number of tweets I was able to work with limited my ability to draw very significant conclusions from the discussion on exonyms, and overall the social media analysis would be made more robust with a larger corpus of data. Future similar projects could potentially draw a larger sample size from other social media websites such as Facebook, or conduct extensive surveys throughout Serbia and Croatia. It was also counter-intuitive that Croatian Twitter users generated proportionally fewer tweets about language, despite language appearing to be a more important and polarizing topic. It may be easier to reconcile this with more context about how Serbs and Croats interact with social media, so there is ample potential for a deeper dive.

Having limited time, I only conducted 48 street interviews in Zagreb and Belgrade, which were useful for corroborating my analytical findings and contextualizing my claims in the field research chapter. However, it would be helpful to conduct even more interviews in order to further expand on my findings and draw meaningful data from them.

It would also be worth discussing the difference between *de jure* and *de facto* language policy. The recommendations of the language institutes in Serbia and Croatia both largely rely on the media and government listening to their *de facto* policies, but some kind of examination of actual laws in place or the *de facto* “culture of correction” in Croatia may serve to further nuance literature on their effects on identity. This may also apply well to research on other ethnolinguistic groups where language serves as both a uniting and dividing factor. Since Bosnia

and Montenegro were outside the scope of this thesis but are going through a similarly interesting inflection point of language policy, research into language identity in those countries would be fascinating. Although the lack of a real name for the shared language in the former Yugoslav space is somewhat unique, similar studies can also be done with other populations which straddle the ambiguous line between dialect and language. A study with an international relations lens, for example, could examine the impact of bilateral ties or international presence of countries on whether systems of communications get classified as dialects or languages.

Future linguistic research on the effects of language policy on identity may also choose to focus on the Sapir-Whorf hypothesis, which didn't receive much discussion beyond the literature review of this thesis. There may be merit in using language relativity and the link between language and thought as perspectives from which to discuss government behavior toward language. Other linguistic angles which were outside the scope of this thesis include how Serbian and Croatian are thought about by diasporic communities, how foreign languages like English are thought about in Serbia and Croatia, and how bilingual speakers think about their language. At the very least, these are all interesting starting points that future work may wish to build on.

Language is fundamental to all of our thoughts and interactions, and affects who we are in so many ways that are often overlooked. In Serbia and Croatia, however, language goes beyond simply affecting identity, and is perhaps the most important factor in defining nationhood. From Serbo-Croatian to Serbian

and Croatian, the difference between a dialect and a language made up the difference between being one country and being two countries.

Appendix A: Sentiment Analysis Code

The following code is a Python script that I ran on a website called Replit and linked to my OpenAI account. Through it, I was able to input a csv file of Tweets, which were processed through GPT-3 to produce a translation, explanation, and sentiment analysis based on the explanation, and output that as another spreadsheet:

```
import os
import openai
import csv
import re

openai.api_key = os.environ['OPENAI_API_KEY']

spreadsheet_file_name = "srpski-serbia.csv"

tweet_column = 1 # note 0 == first column

num_tweets_to_process = None # None = do all

prompt = """"PROMPT""""

def get_completion(tweet_text):
    prompt_text = prompt + tweet_text.replace("\n", "") + "\nTranslation:"
    result = openai.Completion.create(model="text-davinci-002",
                                     prompt=prompt_text,
                                     max_tokens=400,
                                     temperature=0.7)
    return result["choices"][0]["text"]

with open(spreadsheet_file_name, 'r') as f:
    reader = csv.reader(f)
    tweets = list(reader)

tweets_to_process = tweets[
    1:1 + num_tweets_to_process] if num_tweets_to_process else tweets[1:]
output_file_name = spreadsheet_file_name.split('.')[0] + "_result" + '.csv'
# final_output = []
```

```

with open(output_file_name, 'w') as f:
    writer = csv.writer(f)
    writer.writerow([
        "tweet",
        # "raw_output",
        "translation",
        "explanation",
        "sentiment"
    ])

for idx, tweet in enumerate(tweets_to_process):
    tweet_text = tweet[tweet_column].replace("\n", "")
    # print("\n---\n")
    print(f'Analyzing Tweet {idx}')
    # print(tweet_text)
    res = get_completion(tweet_text)
    # print("\n---\n")
    # print("Raw GPT-3 Output:")
    # print(res)
    # print("\n---\n")
    try:
        translation = re.search(r'(.*)', res).group(1)
    except Exception:
        translation = "[not found]"
    try:
        explanation = re.search(r'Explanation: (.*)', res).group(1)
    except Exception:
        explanation = "[not found]"
    try:
        sentiment = re.search(r'Sentiment: (.*)', res).group(1)
    except Exception:
        maybe_positive = "positive" in explanation.lower() and not "not positive" in
explanation.lower()
        maybe_negative = "negative" in explanation.lower() and not "not negative" in
explanation.lower()
        if maybe_positive and maybe_negative:
            sentiment = "[not found]"
        elif maybe_positive:
            sentiment = "Positive?"
        elif maybe_negative:
            sentiment = "Negative?"
        else:
            sentiment = "[not found]"
    # print(
    #     "Parsed:", {
    #         "translation": translation,

```

```
# "explanation": explanation,
# "sentiment": sentiment,
# })
# final_output.append({
# "tweet": tweet_text,
# # "raw_output": res,
# "translation": translation,
# "explanation": explanation,
# "sentiment": sentiment,
# })

# for row in final_output:
# writer.writerow([
# row["tweet:"],
# # row["raw_output"],
# row["translation"],
# row["explanation"],
# row["sentiment"]
# ])

writer.writerow([
tweet_text,
translation,
explanation,
sentiment,
])
```

Appendix B: GPT-3 Prompts

The examples input into GPT-3 were important because they affected how the AI generated its own translations, explanations, and sentiment analysis. I tried to select the example tweets and write my own explanations as neutrally as possible, but am including them here in the interest of full disclosure. These is what I input instead of ““PROMPT”” when working on the code for my sentiment analysis of people using the phrase *srpski jezik*:

Srpski jezik

""Tweets from Serbia and whether their sentiment towards the Serbian language is extremely positive, positive, negative, or neutral/unclear. Extremely positive sentiments actively argue that the Serbian language is its own language, independent of the Croatian language, the Serbo-Croatian language, or the Yugoslav language. Positive sentiments somehow acknowledge the existence of Serbian as a language without directly advocating for it. Negative sentiments somehow criticize or deride the idea of Serbian being an independent language or a language people should learn.

===

Tweet: @radecubura Da, i ja govorim Srpski jezik i ravno pravno koristim oba pisma.

Translation: Yes, I speak the Serbian language and I equally use both scripts.

Explanation: By claiming to speak the Serbian language, the author implicitly acknowledges its existence as an independent language, making this a positive sentiment.

Sentiment: Positive

===

Tweet: Dida je rekao: „A treće i treće, nema bolje stvari nego kad se izmišaju hrvacki i srpski jezik! Tu nastane čisti seks!“ Ja sam pitao: „Koji seks, dida?“ Dida je rekao: „Srpskohrvacki! Tojest hrvackosrpski, zavisi sa koje strane svita gledaš! Dva jezika, isti kurac!“

Translation: Dida said: "And the third and the third, there's nothing better than when the Croatian and Serbian languages are invented! That's when pure sex happens!" I asked: "What sex, grandpa?" Dida said: "Serbo-Croatian! That is, Croatian-Serbian, depending on which side of the world you're looking at it! Two languages, same dick!"

Explanation: The author claims that the Serbian language is invented, making this a negative statement.

Sentiment: Negative

===

Tweet: @dzadzevic @milijanamikicaa @kisjuhas A ti, prije toga, zaboga, nauci srpski jezik! Bar toliko, ako vec hoces da se raspravljias! Eto, napisala sam latinicom, da se vi, velikoumni "BeograDZani" ne smarate sa cirilicom!

Translation: @dzadzevic @milijanamikicaa @kisjuhas And you, before that, for god's sake, learn the Serbian language! At least that much, if you're going to argue! Here, I wrote in Latin, so you, great and wise "Belgradians" don't get confused with Cyrillic!

Explanation: By urging others to learn Serbian, the author nationalistically advocates for a separate Serbian language, making this an extremely positive sentiment.

Sentiment: Extremely Positive

===

Tweet: Aca Seltik na plakatima za akciju "negujemo srpski jezik"??? Koji kralj je to smislio...

Translation: Is Aca Seltik on posters for the action "we nurture the Serbian language"??? What king came up with that...

Explanation: The author implies that the Serbian language does not need to be nurtured, making this a negative sentiment.

Sentiment: Negative

===

Tweet: Srpski jezik je stvarno živopisan ali ne postoji reč za ovo 🙌

<https://t.co/zGtAUsFig4>

Translation: The Serbian language is really colorful but there is no word for this 🙌 <https://t.co/zGtAUsFig4>

Explanation: By mentioning the Serbian language, the author implicitly acknowledges its existence, making this a positive sentiment.

Sentiment: Positive

===

Tweet: Ali OPATICA odo da umrem jebem ti srpski jezik

Translation: But OPATICA I want to die fuck you Serbian language

Explanation: The author expresses frustration with the Serbian language, making this a negative sentiment.

Sentiment: Negative

===

Tweet: Da mi se prevede 'dodjite da nam pravite palacinke' sa kosarkaskog na domaci, srpski jezik, molim vas

Translation: Can someone please translate 'come make pancakes with us' from basketball into everyday, Serbian language

Explanation: By asking for something to be phrased in everyday Serbian, the author acknowledges the existence of Serbian, making this a positive sentiment.

Sentiment: Positive

===

Tweet: Pod hitno uvesti srpski jezik na sve fakultete! Evo #studenti idu ulicama i viču "Dučiću, lopove!". Pojma nemaju.

Translation: Urgently introduce the Serbian language at all universities! Here #students are going through the streets and shouting "Dučiću, robbers!". They don't have a clue.

Explanation: The author is advocating for the use of the Serbian language, making this an extremely positive sentiment.

Sentiment: Extremely Positive

===

Tweet: Novo srpski jezik na delu u dnevnom listu Alo <http://t.co/VudvQfu>

Translation: New Serbian language in action in the daily newspaper Alo <http://t.co/VudvQfu>

Explanation: The author is acknowledging the existence of the Serbian language, making this a positive sentiment.

Sentiment: Positive

===

Tweet: """"

Hrvatski jezik

The following is the text I input when creating a sentiment analysis of the phrase *hrvatski jezik*:

""""Tweets from Croatia and whether their sentiment towards the Croatian language is extremely positive, positive, negative, or neutral/unclear. Extremely positive sentiments actively argue that the Croatian language is its own language, independent of the Serbian language, the Serbo-Croatian language, or the Yugoslav language. Positive sentiments somehow acknowledge the existence of Croatian as a language without directly advocating for it. Negative sentiments somehow criticize or deride the idea of Croatian being an independent language or a language people should learn.

===

Tweet: @Hrvpravdom naučite hrvatski jezik Hrvatine!

Translation: Learn the Croatian language, Croatians!

Explanation: The author is urging other Croatians to learn Croatian. This is actively arguing that the Croatian language is a separate language that people should learn, so it is extremely positive.

Sentiment: Extremely positive

===

Tweet: Srpski jezik, Hrvatski jezik, Bošnjački jezik, Crnogorski jezik, sve neki različiti jezici, a gramatika identična u najsitnijim detaljima. Mnogo smo bliži nego što to neki hoće da prikažu.

Translation: Serbian language, Croatian language, Bosnian language, Montenegrin language, all kinds of different languages, but the grammar is identical in the smallest details. We are much closer than some want to show.

Explanation: The author believes these languages are all extremely similar. That indicates they don't think of Croatian as a separate language, which is a negative opinion of the language.

Sentiment: Negative

===

Tweet: Hrvatski jezik 'udario je kameno dno' s trendom doslovnog prijevoda engleskih fraza na hrvatski. I osobno sam zaslužan za mali dio tog trenda, a primjera ima hrpa i po lajni. 😬 <https://t.co/3O1aYGpZzQ>

Translation: The Croatian language 'hit rock bottom' with the trend of literal translations of English phrases into Croatian. And I'm personally responsible for a small part of that trend, and there are plenty of examples in my timeline. 😬

<https://t.co/3O1aYGpZzQ>

Explanation: Although the author criticizes the practice of literally translating English phrases, they nevertheless passively acknowledge its existence as an independent language, making this a positive opinion of the existence of Croatian.

Sentiment: Positive

===

Tweet: Upravo sam naučio da hrvatski jezik ne razlikuje u prijevodu "channel" i "canal", a pojmovi nisu potpuni sinonimi.

Translation: I just learned that the Croatian language does not differentiate in the translation of "channel" and "canal", and the concepts are not complete synonyms.

Explanation: By mentioning the Croatian language, the author passively acknowledges its existence, making this a positive sentiment.

Sentiment: Positive

===

Tweet: @dnolegy Oni imaju lektore? Nemreš bolivit. Sigurno su kratkovidni jadni, a i možda jesu lektori ali ne za hrvatski jezik?

Translation: @dnolegy Do they have lecturers? You can't change them. They're probably short-sighted, poor, and maybe they are lecturers but not for the Croatian language?

Explanation: Despite criticizing the lecturers, this tweet refers to the Croatian language in a manner that acknowledges its existence, making this a positive sentiment.

Sentiment: Positive

===

Tweet: @gogePN Hrvatski jezik je presiromašan za točno opisati što se tu dogodilo 😄😄😄

Translation: The Croatian language is too poor to accurately describe what happened here 😄😄😄

Explanation: Although it laments the shortfalls of the Croatian language, this tweet nevertheless characterizes it as a language, implicitly different from the Serbian language, so this is a positive sentiment.

Sentiment: Positive

===

Tweet: 1847. na posljednjem zasjedanju Hrvatskog sabora nakon prijedloga I.K.Sakcinskog, hrvatski jezik je proglašen službenim u javnoj uporabi.

Translation: 1847. at the last meeting of the Croatian Parliament, after the proposal of I.K.Sakcinskog, the Croatian language was declared the official language in public use.

Explanation: The tweet is stating a historical fact and does not reveal anything about the opinion of the author.

Sentiment: Neutral/Unclear

====

"Hitne službe nije dovoljno dobar termin za hrvatski jezik? Užas

Translation: "Emergency services" is not a good enough term for the Croatian language? Disgusting

Explanation: The author is criticizing the term "emergency services" as being insufficient for the Croatian language. This is criticizing Croatian language policy, making this a negative sentiment.

Sentiment: Negative

====

Tweet: I got a 90 (devedeset) on my mid-class language exam (hrvatski jezik ispit) let's fucking go! (Jebo-something Idemo!)

Translation: I got a 90 (devedeset) on my mid-class language exam (hrvatski jezik ispit) let's fucking go! (Jebo-something Idemo!)

Explanation: The author is celebrating a high score on their Croatian language exam. This is a positive sentiment because it acknowledges Croatian as its own language.

Sentiment: Positive

====

Tweet: Plenković: Sramotno negiranje hrvatskog jezika u srbijanskom udžbeniku <https://t.co/RR2DqUCRbk> <https://t.co/xA7llfTZUM>

Translation: Plenković: The disgraceful denial of the Croatian language in the Serbian textbook <https://t.co/RR2DqUCRbk> <https://t.co/xA7llfTZUM>

Explanation: The author is criticizing the Serbian textbook for denying the existence of the Croatian language. This is an extremely positive sentiment because it actively advocates for acceptance of the Croatian language.

Sentiment: Extremely positive

====

Tweet: """"

Naš jezik

The following is the text I input when creating a sentiment analysis of the phrase *naš jezik*:

""""Recently, a movement has emerged in Serbia and Croatia to not refer to their highly politicized shared language as "Serbian" or "Croatian," but instead call it "our" language to present a unified front.

Analyze these tweets from Serbia and whether their sentiment towards a unified "our" language is extremely positive, positive, negative, or neutral/unclear. Extremely positive sentiments actively argue that Serbian and Croatian are one language that should be called "our" language, or somehow advocate for the idea of calling the language by a single name. Positive sentiments somehow acknowledge the existence of "our language" as a term without directly advocating for or against it. Negative sentiments somehow criticize or deride the idea of a unified language, or only use "our" to refer to the Serbian or Croatian language.

===

Tweet: @findit033 @Dalmacija_1995 @Rillyricum posrbļjava jezik bastini i naslidjuje od nas. Mi smo prvi. Serbo je samo adoptira nas jezik.

Translation: he uses the language of heritage and imitates it from us. We are the first. Serbian is just an adoption of our language.

Explanation: This is a negative sentiment because it derides the idea of Serbian and Croatian being unified, only using the phrase "our" to refer to the Croatian language.

Sentiment: negative

===

Tweet: @swinkisonfire ah jebiga, ne pratim nju. :D Ritvitćem očito prijevode na naš jezik :D cc @leatherpleasure

Translation: @swinkisonfire ah man, I don't follow her. :D Ritvitć obviously translates to our language :D cc @leatherpleasure

Explanation: This tweet is positive because, by referring to "our" language, it implicitly acknowledges the idea of a unified language and espouses the movement to refer to it as "our."

Sentiment: positive

===

Tweet: Čitam danas tako o hrvatskom jeziku i razmišļjam koliko se isti svakim danom, sve više i više zanemaruje. A zaista je lijep taj naš jezik.

Translation: I'm reading today about the Croatian language and thinking about how much it's neglected every day, more and more. And our language is really beautiful.

Explanation: This tweet has a negative sentiment because in this sentence "our" refers to the Croatian language, which means the author thinks it is a separate language.

Sentiment: negative

===

Tweet: "@DarkoLesinger Cijepljeno , procijepljeno , docijepljeno , nacijepljeno , pricijepljeno ... LIJEPI NAS JEZIK"

Translation: "@DarkoLesinger Vaccinated, revaccinated, boosted, nationalized, internationalized ... OUR LANGUAGE IS BEAUTIFUL"

Explanation: This tweet has a positive sentiment because it refers to "our" language without explicitly advocating for a unified language

Sentiment: positive

===

Tweet: @AntoniaLG1 ima susjedu Rumunjku koja je negdje prepoznala naš jezik, prišla ženi, dala joj njen broj i evo idu dvije Hrvatice na babinjak. ❤️

Translation: @AntoniaLG1 she has a Romanian neighbor who somewhere recognized our language, approached the woman, gave her her number and now two Croatian women are going to grandma's. ❤️

Explanation: This tweet is positive because, by referring to "our" language, it implicitly acknowledges the idea of a unified language and espouses the movement to refer to it as "our."

Sentiment: positive

===

Tweet: ""

Appendix C: Sample Interview Guide

Here are the interview questions I asked my randomly sampled respondents, including translations into English:

1. Odakle ste?

Translation: where are you from?

2. Kako se zove jezik koji govorite?

Translation: what's the name of the language you speak?

3. Kako bi ste opisali razliku izmedju srpskog i hrvatskog?

Translation: how would you describe the difference between Serbian and Croatian?

4. Šta je vaše mišljenje o upotrebi izraza “naš jezik” za oba srpski i hrvatski?

Translation: what are your thoughts on the use of the phrase “naš jezik” for both Serbian and Croatian?

5. Šta utiče na vaše mišljenje o vašem jeziku?

Translation: what influences how you think about your language?

6. Kako je politika i raspad jugoslavije uticalo na vaš jezik?

Translation: how have politics and the dissolution of Yugoslavia influenced your language?

7. Koliko imate godina?

Translation: How old are you?

The interviews with the academics and policymakers were more conversational and did not follow a set question list.

Bibliography

- Anzaldúa, Gloria. "How to Tame a Wild Tongue." *The Mexico Reader*, Duke University Press, 2022, p. 34, <https://doi.org/10.2307/j.ctv2rr3g8m.12>.
- Bailyn, John Frederick. "To What Degree Are Croatian and Serbian the Same Language?: Evidence from a Translation Study." *Journal of Slavic Linguistics*, vol. 18 no. 2, 2010, p. 181-219. *Project MUSE*, doi:10.1353/jsl.2010.0000.
- Bakić-Hayden, Milica. "Nesting Orientalisms: The Case of Former Yugoslavia." *Slavic Review*, vol. 54, no. 4, 1995, pp. 917–31. *JSTOR*, <https://doi.org/10.2307/2501399>. Accessed 8 Mar. 2023.
- Barić, Eugenija., et al. *Hrvatski jezični savjetnik / c Eugenija Barić ... [et al.] ; u redakciji izvršnog uredništva Lane Hudečak, Milice Mihaljević i Luke Vukojevića*. Institut za hrvatski jezik i jezikoslovlje : Pergamena : Školske novine, 1999.
- Boduszyński, Mieczysław P., and Vjeran Pavlaković. "Cultures of Victory and the Political Consequences of Foundational Legitimacy in Croatia and Kosovo." *Journal of Contemporary History*, vol. 54, no. 4, 2019, pp. 799–824. *JSTOR*, <https://www.jstor.org/stable/26787449>. Accessed 7 Mar. 2023.
- Bugarški, Ranko. "Language, Identity and Borders in the Former Serbo-Croatian Area." *Journal of Multilingual and Multicultural Development*, vol. 33, no. 3, 2012, pp. 219–35, <https://doi.org/10.1080/01434632.2012.663376>.
- Bugarški, Ranko. "Language, Nationalism and War in Yugoslavia." *International Journal of the Sociology of Language*, vol. 2001, no. 151, 2001, pp. 69–87, <https://doi.org/10.1515/ijsl.2001.048>.
- Bugarški, Ranko. "Language Policies in the Successor States of Former Yugoslavia." *Journal of Language and Politics*, vol. 3, no. 2, 2004, pp. 189–207, <https://doi.org/10.1075/jlp.3.2.04bug>.
- DellaVigna, Stefano, et al. "Cross-Border Media and Nationalism: Evidence from Serbian Radio in Croatia." *American Economic Journal. Applied Economics*, vol. 6, no. 3, 2014, pp. 103–32, <https://doi.org/10.1257/app.6.3.103>.
- Dulčić, Mihovil. *Govorimo hrvatski: jezični savjeti*. Hrvatski radio : Naprijed, 1997.

- Greenberg, Robert D. "Language and Identity in the Balkans." *Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration*, Oxford University Press, 2008, <https://doi.org/10.1093/acprof:oso/9780199208753.001.0001>.
- Greenberg, Robert D. "Language, Nationalism and the Yugoslav Successor States." *Language, Ethnicity and the State*, Palgrave Macmillan UK, pp. 17–43, https://doi.org/10.1057/9781403914187_2.
- Laitin, David D. "Politics, Language, and Thought: The Somali Experience." *Politics, Language, and Thought: The Somali Experience*, 1977, pp. xi–268.
- Langston, K., and A. Peti-Stantić. *Language Planning and National Identity in Croatia*. 1st ed., Palgrave Macmillan UK : Imprint: Palgrave Macmillan, 2014.
- Jelić, Magdalena Petrović. "Serbian, Croatian, Bosnian, or Montenegrin: Our Language." *Serbonika*, <https://serbonika.com/blog/serbian-language/serbian-and-croatian/>.
- Kemp, Simon. "Digital 2022: Croatia." *DataReportal*, Global Digital Insights, 15 Feb. 2022, datareportal.com/reports/digital-2022-croatia.
- Kemp, Simon. "Digital 2022: Serbia." *DataReportal*, Global Digital Insights, 15 Feb. 2022, datareportal.com/reports/digital-2022-serbia.
- Komarzyca, Daniel, and Janina Frasz. "Language and Politics in India and China: A Cross-Cultural Comparative Study." *Polish Political Science*, vol. 49, no. 4, 2020, pp. 9–36, <https://doi.org/10.15804/ppsy2020402>.
- Makoni, Sinfrey, and Alastair Pennycook. "Disinventing and Reconstituting Languages." *Disinventing and Reconstituting Languages, Multilingual Matters*, 2018, pp. 1–41, <https://doi.org/10.21832/9781853599255-003>.
- O'Reilly, Camille. *Language, Ethnicity, and the State*. Palgrave, vol. 1, 2001.
- Pupavac, Vanessa. *Politics and Language Rights: A Case Study of Language Politics in Croatia*. Springer Nature, 2003.
- Rice, Eric A. *Language Politics in Bosnia, Croatia, and Serbia*. 2010.
- Stanković, Branimir, and Marija Stefanović. "Peeling the Onion Top-down: Language Policy in Serbia Between Power and Myth." *Aegean Working Papers in Ethnographic Linguistics*, vol. 2, no. 1, 2019, <https://doi.org/10.12681/awpel.20022>.

Stojiljković, Vukašin. "Negotiating the Yat Border(s)." *Wiener Linguistische Gazette*, no. 81, 2017, p. 47.

Turk, Marija, and Maja Opašić "Linguistic Borrowing and Purism in the Croatian Language." *Suvremena Lingvistika*, vol. 34, no. 65, 2008, pp. 73–88.