9-2022

Predicting Stress in Russian using Modern Machine-Learning Tools

John Schriner
The Graduate Center, City University of New York

PREDICTING STRESS IN RUSSIAN USING MODERN MACHINE-LEARNING TOOLS

by

JOHN SCHRINER

A master's thesis submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Master of Arts, The City University of New York

2022

Predicting Stress in Russian using Modern Machine-Learning Tools

by

John Schriner

This manuscript has been read and accepted for the Graduate Faculty in Linguistics in satisfaction of the thesis requirement for the degree of Master of Arts.

__________________	________________________________
Date	Kyle Gorman Thesis Advisor
__________________	________________________________
Date	Cecelia Cutler Executive Officer

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

PREDICTING STRESS IN RUSSIAN USING MODERN MACHINE-LEARNING TOOLS

by

JOHN SCHRINER

In the Russian language, stress on a word is determined via often complex patterns and rules. In this paper, after examining nearly a century of research in stress rules and methods in Russian, we turn to see if modern machine learning tools can aid in predicting stress. Using A.A. Zaliznyak's dictionary grammar and over 300,000 word forms, we derived stress codes to aid in predicting which syllable primary stress falls on. We trained an LSTM neural network on the data and conducted eight experiments with added features such as lemma, part of speech, and morphology. While the model performed better than baseline in most experiments, the lemma feature outperformed every other feature.

ПРОГНОЗИРОВАНИЕ СТРЕССА В РУССКОМ ЯЗЫКЕ С ИСПОЛЬЗОВАНИЕМ СОВРЕМЕННЫХ СРЕДСТВ МАШИННОГО ОБУЧЕНИЯ

В русском языке ударение в слове часто определяется сложными схемами и правилами. Опираясь на почти сто лет исследований правил ударения в русском языке, в этой статье мы обращаемся к вопросу, могут ли современные инструменты машинного обучения помочь в предсказании ударения в слове. Используя грамматический словарь А.А. Зализняка и более 300,000 словоформ, мы вывели коды, которые помогают определить, на какой слог будет падать основное ударение. Мы обучили нейронную сеть LSTM на этих данных и провели восемь экспериментов с моделью, добавив дополнительные параметры:

лемма, часть речи и морфология. В большинстве экспериментов эта модель работала лучше, чем базовая, а параметр леммы превзошел все остальные параметр.

ACKNOWLEDGMENTS

I wish to thank first and foremost Dr. Kyle Gorman who helped me to shape and fine-tune this project from the start. I am very appreciative of all of the work in this area that came before, as well as the Zaliznyak data thoroughly documented and provided by the Giella project and Yulia Spektor. Thank you to all of my colleagues in Linguistics at the CUNY Graduate Center for being endlessly curious and inspiring about linguistics and computational linguistics in particular. Thank you to Darya Bukhtoyarova also for linguistic inspiration and for bouncing ideas back and forth over the years. A heartfelt thanks to Andrea for your vast support in this and everything.

Contents

List of Figures

List of Tables

1. Introduction

To pronounce Russian words correctly, the speaker needs to know where to place the primary stress. Experiments like the ones below are therefore relevant for text-to-speech tasks, automatic speech recognition, computer-assisted language-learning, and ultimately furthering our understanding of the underlying patterns for the placement of stress in the Russian language.

The inquiry into rules and predictable patterns for Russian stress has been the serious focus of researchers for over a century. While several suggest that Russian has unpredictable stress, others maintain that rules exist that can be codified. In 1977, A.A. Zaliznyak furthered this endeavor by authoring a grammar and dictionary that contained codes for stress.

Stress is not marked orthographically in Russian with these few exceptions: to remove ambiguity; to note pronunciation, perhaps for archaic words; to note pronunciation of certain professional non-Russian words; words specific to a dialect or slang; in poetic verse (Wade & Gillespie, 2011:§1.15). Wade & Gillespie (2011) explain that stress in Russian is free, and a "change in stress may indicate a change in meaning" (§1.15) using the example of о́рган 'organ of the body' and орга́н 'organ' (musical instrument). Incorrectly stressed words may be unintelligible. Lavitskaya & Kabak (2014) maintain that "stress assignment is largely unpredictable, suggesting that word stress must be stored in the mental lexicon" (p. 363).

In this paper, when referring to Russian words or affixes where the gloss in English is desirable, the affix will be followed by the gloss in single quotes: e.g. -изм '-ism'. Particular Cyrillic letters will be noted with guillemets, e.g. «ы». Rather than transliterating into Latin script, to avoid ambiguity, original Cyrillic script will be used throughout this paper.

1.1 Uncovering the Patterns

"Stress may be described as the degree of force with which a sound or syllable is uttered" (Jones, 1956 as cited in Ward, 1965, p. 45). "Russian stress is 'free' and 'mobile' and is closely associated with changes in vowel quality" (Ward, 1965, p. 45). Ward (1965) continues: "a stressed vowel in Russian is slightly longer than an unstressed vowel" (p. 46). Russian stress is free in that it can occur on any syllable and "is not restricted to, say the initial syllable, as it is in Czech, or the penultimate, as it is almost entirely in Polish" (Ward, 1965, p. 47). Stress is mobile in that the stress placement may move to a different vowel in different grammatical forms of the same word. Table 1 shows examples from the literature of mobile stress. While stress is not ordinarily marked orthographically, in data concerning primary stress, it is marked on vowels with an acute accent diacritic. While the use of stress remains constant in Russian, "changes in the location of stress are [...] constantly occurring" (Ward, 1965, p. 45). "There is a steady tendency to regularize the location of stress in some sets of words, yet this very tendency itself creates further 'anomalies'" (Ward, 1965, p. 45).

Table 1 - Examples of Mobile Stress

Водá 'water'	Вóду	(sg/acc. sg)	(Hingley, 1952, p. 186)
Рукá 'hand'	Рýки	(sg/pl)	(Jouravlev & Lupker, 2015, p. 945)
Óблако 'cloud'	Облакá	(sg/pl)	(Ponomareva et al., 2017:§2.2)

Wade & Gillespie (2011), in their Comprehensive Russian Grammar, note that Russian has ten vowel letters that can be classified as back, central, and front. Unstressed «и» and «ы» are shorter and "pronounced in a more 'relaxed' fashion" (Wade & Gillespie, 2011:§1.4), while «о» and «а» are similarly reduced when unstressed. The reduction and exact pronunciation of «е» and «я» depends on whether the letter is pre-tonic or post-tonic.

The vowel «ё» does not appear in unstressed position; that is to say, «ё» is always stressed. While older sources, in particular, add primary or secondary stresses elsewhere in words, this paper will maintain that when «ё» occurs, it will bear primary stress.

Jakobson (1957) notes that "Russian declension exhibits a clear tendency to express [the] distinction between [the case and number markers] by a mere difference in the place of wordstress" (p. 184). Jakobson (1957) also notes the *—a* declension as an example of a shift in stress from the desinence to the stem as "an adaptation to the pattern of nouns with mobile stress" (p. 185). Such drift can be found by close observation of changing forms over time.

While we might agree with Shapiro (1986) that "Russian stress is a puzzle" (p. 183), it is not, however, capricious. Shapiro (1986) provides two important points: firstly, "Russian stress is part of the morphology and morphophonemics of Russian and basically not motivated by the phonology (while having phonological consequences, namely vowel reduction)" (p. 183). Secondly, "though stress is in principle free to occur on any syllable of a word, certain restrictions are known to obtain whereby the morphophonemic or morphological structure of a given word, or its membership in a particular form class or semantic category, narrow the range of possible accent positions" (Shapiro, 1986, p. 183). The complexity may mean that "at first sight it would not appear that there are many 'rules' (for that is what the correlation of stress and word-type or grammatical form amounts to) for the location of stress" (Ward, 1965, p. 48). Hart (2015) recalls that Zaliznyak was "not sanguine about the possibility of developing a single fundamental rule for locating stress in any given word" (p. 2). Indeed, due to the complexities of inflectional and derivational morphology, "any attempt to formulate it would be pointless (нецелесообразно)" (Hart, 2015, p. 2) or impractical. Ward (1965) delineates some of the rules at random: the suffix -ция '-tion' always has stress on the preceding syllable; the suffixes -изм '-ism' and -ист '-ist' always have the stress; the suffix -олог '-ologist' receives the stress on the first «о» while -ология '-ology' receives the stress on the second «о»; the adjectival suffix -ический '-al, -ic, -ical' is always stressed on the first «и»; perfective verbs with the prefix вы- 'out, to completion' always receive the stress on the «ы» (pp. 48-49).

1.2 Paradigms and Subparagdigms

Feldstein (1980; 1986; 1993; 1996) thoroughly developed the theoretical foundation of a paradigmatic method of classifying Russian stress patterns and its inception was first attributed to Leonard Bloomfield in 1945. Each grammatical category may be conceptualized as an accentual paradigm: while nouns have subparadigms of singular and plural, verbs have subparadigms of present and past, and adjectives have subparadigms of long-form (attributive) and short-form (non-attributive) (Feldstein, 1980). A two-letter index was devised and each part-of-speech "contains exactly the same inventory of stress types at a deeper than surface structural level, although.. [they] will vary predictably in the different morphological categories" (Feldstein, 1980, p. 124). Feldstein (1993) suggests using a two-stage procedure that first establishes each individual lexeme's accentual pattern, and then secondly grouping each of these patterns by stem-stress, ending-stress, and mobile stress. Feldstein (1993) observes that this grouping is "not at all an obvious procedure" (p. 47) due to historically different approaches to grouping that lead to different results. Zaliznyak also understood the importance of establishing subparadigms that were essentially morphological paradigms divided into parts. Semantic features of a word that affect stress include abstractness, countability, and for adjectives, qualitative or relational features that tend to link it to a stress type. Further, "Zaliznyak's correlation of certain semantic features with accentual features has led to a scheme in which certain lexical features are held constant so that particular lexical classes can be singled out for their accentual behavior" (Feldstein, 1993, p. 57). Feldstein (1993) concludes that our understanding of subparadigms best falls on a continuum that each serve different analytical purposes "depending on whether these accentual, segmental, and semantic features are taken as constant or variable within the paradigm under consideration" (p. 59).

Halle (1973) postulates several rules that account for stress. Given that a word has a [+Stress] and is theoretically marked on a vowel, he proposes an S-Distribution rule that gives each vowel preceding the stressed vowel a stress marker as well. The Destressing rule will eventually remove all stresses except for the last stressed vowel. The Circumflex rule assigns stress to the initial syllable. Stressless vowels undergo neutralization, which "consists of a merger of all non-high vowels into a single vowel" (Halle, 1973, p. 313). When a potentially stress vowel is deleted, the stress usually moves to the preceding vowel. The order of the rules is important: any deletion necessarily comes before the Destressing rule. Halle (1973) maintains that deletion may result in a stressless word. Echoing Jacobson (1948), Halle notes that Russian words are subject to the Vowel Truncation rule that deletes a morpheme-final vowel in position before a vowel. The Yers rule states that when the back yer «ъ» or the front yer «ь» are placed before another yer, they are converted to «e» or «o» or deleted entirely. Interestingly, Halle (1973) notes that these rules track across different dialects, not just the literary form.

Nouns have perhaps the simplest pattern, accounting for 90 percent of the nouns in the Russian dictionary: stress is placed on the same vowel of the stem in all case forms (Halle, 1973). The exception to this is when a yer is present and deleted.

Coats (1976) details four distinct stress types for nouns: stem-stressed type, in which "the stem is stressed in all forms of the paradigm"; the end-stressed type, in which the "endings are stressed in all forms of the paradigm"; the mobile-stress type, in which the "stress falls on the ending in some forms of the paradigm and on the *initial* syllable of the stem in others"; and finally the retracted-stress type, in which "stress falls on the ending in some forms and on the *final* syllable of the stem in others" (p. 1, emphasis in original).

Coats (1976) notes that stress likely cannot be predicted in stem-stressed words even by knowing the word's "phonological, morphological, syntactic, or semantic properties" (p. 2) and that stress ought to be marked on the root in the lexicon. Coats describes the work of Halle (1973), noting that end-stressed words may be analyzed as conforming to the *oxytone rule* whereby words with no stressed root will be assigned stress on the endings. Coats, instead, developed his Stress Adjustment Rule that, on the surface seems similar, but in fact can account for end-stressed words as well as verbs with inherently stressed suffixes.

According to Coats (1976), mobile-stressed words are "accentually irregular" as they are "subject to Stress Deletion and Initial Stress Rules in certain paradigmatic forms only" (p. 6). These rules may act in concert or independently throughout Coats' analysis. They show how some forms have an initial stress while some are assigned to the last syllable. Coats organized his work by part of speech and delineated rules and exceptions to the rules. Similar to mobile-stress and end-stress, retracted-stress words are said to have unstressed stems, but in this case they undergo the Stress Retraction Rule that retracts stress from the last syllable to the preceding syllable, again, in certain paradigmatic forms that contain morphemes that are susceptible to the rule (Coats, 1976).

Hingley (1952) works to describe the "linguistic perversities" of stress in Russian: "in no grammatical category do the fluctuations of stress under inflection (‘flexional stress’) present a greater appearance of chaos than among nouns in -а/-я” (p. 186). Hingley (1952) showed that nouns of a basic final stress are in a state of flex, changing as languages generally do; this can be seen by identifying forms with hesitant stress, e.g. соха́ ‘plow’ (со́ху́; со́хи; со́ха́м).

Lavitskaya & Kabak (2014) initiated two production experiments aiming to reveal stress patterns in nouns: “final stress in consonant-final words, penultimate stress in vowel-final words” (p. 379). Their first experiment used fictitious place names that ended in —*i* or —*o*. This resulted in: “all of our vowel-final words were not declined by the participants, whereas all consonant-final words were declined” (Lavitskaya & Kabak, 2014, p. 375), i.e. an inflectional ending was added by the participants. Their second experiment used novel acronyms in an effort to reveal default stress placement. Acronyms in this experiment were penultimately stressed in about 78% of the cases and suggested that stress distribution in the stimuli depend on the quality of the vowel (Lavitskaya & Kabak, 2014).

Hart (2015) focuses on Russian derived nouns and finds that “stress in Russian derived nouns functions to define the boundary between the root and the suffix” (p. 1). Further, he finds that “the presence of a phonological factor in what traditionally has been seen as a strictly morphological phenomenon in the modern language is indicated” (Hart, 2015, p. 1). By examining the stress characteristics of suffixes, Hart (2015) hypothesizes that the ultima, or stem-final syllable, plays an important role in the location of stress. To find evidence for a default stress location, Hart (2015) investigates the suffixes -ик (masculine) and -ица (feminine) for nouns, which account for over 5000 words in Zalyznyak’s (1977) dictionary. No nouns with these suffixes have mobile stress, so the stress remains on the same vowel throughout the paradigm and is either a fixed root stress, a fixed stem stress, or a fixed end stress. In the fixed stem stress pattern (i.e. having stress on the suffix), stress is placed on the first vowel of the suffix if it is vowel-initial; if the suffix begins with a consonant, then stress is placed on the first vowel that precedes the suffix. Compound suffixes make descriptive work difficult as they may be analyzed with ambiguous constituent structures. An addition of a suffix to a root can create a new constituent, or *fused root* (Hart, 2015). Hart (2015) noted that 82% of all words with the -ица suffix had stem stress: "in these words, stress appears adjacent to the suffix on the root portion of the stem marking the intersection of root and suffix" (p. 9). Hart's (2015) research furthers the idea that by knowing the boundary of the root and the suffix, it may be clearer where to place stress.

There has been no consensus about the stress patterns of adjectives in Russian: while Kempgen (1990) states "for Russian adjectives it is well known that there exists a tendency to place the accent near the end of the word" (p. 210), Jouravlev & Lupker (2014) propose that "Russian adjectives seem to be much more likely to have first-syllable stress" (p. 606). In their experiments, described at length below, Jouravlev & Lupker (2014) find that in Russian disyllabic words, adjectives have trochaic stress 80% of the time.

1.3 Coding Systems

Several prominent coding systems in grammars were devised to provide a word's paradigmatic properties. This paper focuses on data derived from Zaliznyak's coding in his dictionary grammar. This work "contains a definitive description of Standard Russian inflectional patterns, still unsurpassed, due to its depth and precision" (Iosad et al., 2018, para. 2). His dictionary grammar is presently used as a tool for natural language processing, and is "all the more impressive by the fact that Zaliznyak did all the work by hand, collating the data on slips of paper" (Iosad et al., 2018, para. 2). To use Zaliznyak's coding, one would first use the direct method to consult the grammar tables, eventually moving to the analytic method thereby *reading* the codes without having to consult the reference tables (Haraldsson, 1996). Clearly a desirable goal for coding systems is "maximal coordination and standardization of the models for dictionary grammar" (Haraldsson, 1996, p. 562).

1.4 Frequency and Familiarity

"Familiarity is correlated with a mobile stress pattern, unfamiliarity with a fixed pattern–regardless of gender" (Shapiro, 1986, p. 187). A given word may be habitual or alien/unfamiliar on a continuum; habitual items have mobile stress whereas alien words have fixed stress. Words out of fashion, like alien words, tend to have fixed stem stress throughout the paradigm. Among non-professionals, Zaliznyak notes the following standard stress: špríc 'syringe', massáž 'massage', trjúm 'hold of a ship', and bócman 'boatswain' are fixed on the stem; whereas to professionals who use these words every day, the stress pattern is mobile: špricý, massaží, trjumá, bocmaná (Shapiro, 1986, p. 184). Lagerberg (2007) continues connecting frequency to stress patterns: frequency "goes a significant way towards explaining why stress variation and mobility continue to exist, and even flourish, in modern Russian" (p. 166).

Much of Zaliznyak's work deals with conceptualizing the correlation between word frequency and stress patterns: "In the noun, more familiar words tend to be stem-stressed, while less familiar words tend to be ending-stressed in the plural" (Feldstein, 1993, p. 57).

Mustajoki (2008) opines that by "using word frequency tables, it should be possible to resolve the problem of whether the frequency of a substantive in Russian can be taken as an indicator of mobile stress" (p. 200).

1.5 Stress Homographs

Stress homographs exist in Russian and further complicate how one pronounces certain words. Derbyshire (1966) in his research into verbal homonymy explains that "homographs are words with identical orthography, but whose stress and pronunciation are not the same" (p. 131). Lagerberg (2007) provides examples of stress homographs in discussing semantic cases of stress variation: а́тлас 'atlas' and атла́с 'satin'. Naturally, this leads to ambiguity for text-to-speech machines as well as for humans. Gorman et al. (2018) work to disambiguate homographs by creating hybrid systems that use both a rules-based disambiguation system alongside a machine learning-based system. Gorman et al. (2018) explain that Google's text-to-speech system differentiates homographs by using classes of rules based on context, or collocation with other words, part of speech tagging, or lastly a default pronunciation when no other rule is chosen. As discussed in detail below, Zaliznyak's dictionary does not contain words in context, although it does provide part of speech tagging.

1.6 Secondary Stress

Ward (1965) maintains that in Russian polysyllabic words, the stress on each syllable crescendos until the primary stress is reached, leading to "an abrupt fall" (p. 46). "Secondary stress is particularly common in words with foreign prefixes (а́нтикоммуи́зм 'anti-communism')" (Wade & Gillespie, 2011:§1.15), as well as in the aforementioned first part of hyphenated compounds. "Generally speaking, the newer a compound word is, the more likely a secondary stress (e.g. ки́носцена́рий 'film script')" (Wade & Gillespie, 2011:§1.15). Secondary stress is often displayed orthographically in the literature with a grave accent, e.g. а̀втостоя́нки 'car parks'.

1.7 Prior Computational Work

Hall & Sproat (2013) approach Russian stress-prediction as a data-driven ranking problem similar to Dou et al.'s (2009) SVM-based ranking approach but with added affix-based features. They used an expanded Zaliznyak grammatical dictionary that yielded over 2 million fully inflected words. While there were no forms present in both the training and test data, there was certainly shared lemmata across the sets. They extracted 248 forms of words that did not have lemmata present in the training set and created a *no shared lemmata* set with which to compare the results. Adding various affix-based feature combinations, Hall & Sproat find significant accuracy in stress prediction. Taking into account the placement of secondary stress, they note a slight drop in performance.

Jouravlev & Lupker (2014, 2015) approach the question of stress-placement in Russian from the field of reading research. "Readers can assign lexical stress by computing it based on stress rules or non-lexical sources of evidence for stress (stress cues) present in the language" (Jouravlev & Lupker, 2015, p. 944). Their research notes that prior research provided many possible cues for computing lexical stress, but it's difficult to "investigate multiple sources of evidence for stress within a single data set" (Jouravlev & Lupker, 2015, p. 945), where cues may or may not be relied on differently by the reader. Using a set of binary logistic regressions with combinations of 11 predictors, they chose a model that had a minimal number of factors but would still wield a high predictive power. In their experiments, they use disyllables only, as a first step towards understanding the mechanisms of stress-assignment. Jouravlev & Lupker (2015) identify a word's morphology as a potential source of evidence for stress patterns but surmise that morphology can't be the primary cue as it's only available for polysyllabic words, i.e. it would not explain how readers assign stress to monomorphemic words. They note the work of Protopapas et al. (2007) on stress in Greek and the need to take morphology into account: “certain inflectional morphemes carry stress assignment information” (p. 717). Further, they note that grammatical category may also be a source of evidence for stress patterns: “Arciuli and Cupples (2006), however, proposed that the relationship between grammatical category and lexical stress might be artifactual. The word’s orthography might be cuing its grammatical category and its lexical stress at the same time, essentially independent of one another” (p. 947). The essential goal of their study was to “determine whether probabilistic associative relations exist between the examined variables and stress patterns in the Russian language” (Jouravlev & Lupker, 2015, p. 948) and whether or not readers of Russian use these associations. Their study was designed to include even sources that have not been theorized to aid in stress placement. They note that morphological information has limited usefulness as the majority of words are composed of stress-ambiguous morphemes. They conclude: “assigning stress in Russian is a process that involves considerably more than retrieving stress information from lexical memory” (Jouravlev & Lupker, 2015, p. 964). Readers have the knowledge of a “regular stress pattern as defined within the word’s grammatical category, and of the consistency with which a word ending maps onto a stress pattern” (Jouravlev & Lupker, 2014, 612).

Reynold & Tyers (2015) begin by noting that Russian word stress is phonemic, and that “many wordforms are distinguished from one another only by stress position” (p. 173). Also they note that “half of the vowel letters in Russian change their pronunciation significantly, depending on their position relative to the stress” (Reynold & Tyers, 2015, p.173). The researchers developed a finite-state transducer sourcing Zaliznyak’s dictionary and grammar that generated all possible morphosyntactic readings of each wordform, as well as a constraint grammar that took syntactic context into account and removed a portion of the readings (Reynold

& Tyers, 2015). Given a particular set of readings, their output offered results such as: *bare*, meaning that if there was more than one reading, stress would not be marked; *safe*, meaning that if each reading shared a stress marker with no exceptions, stress would be marked at that shared location; *randReading*, meaning that, if for example, there were two readings that shared a stress location and one reading that differed, the former would be twice as likely to be chosen; and lastly *freqReading* chose the tag sequence (i.e. morphological features) that is most frequent, e.g. N-F-SG-GEN is more frequent than N-F-SG-DAT in the dataset. The researchers' use of error-tolerance in the form of the *safe* output provides a certain level of assurance in accuracy: "we are able to abstain from marking stress on tokens whose morphosyntactic ambiguity cannot be adequately resolved" (Reynold & Tyers, 2015, p. 178). The researchers note that their gold standard stress corpus (just 7689 tokens) was a prudent addition to their experiment as processing it could provide context. Their results demonstrated that using a constraint dictionary greatly improved their stress-placement task and by using the corpus of running text, they discovered "the importance of context-based disambiguation for this task" (Reynold & Tyers, 2015, p. 179).

The Russian language has been the focus of many studies involving machine learning due to its grammatical and inflectional complexity, and we thought it important to look at prior sequence-to-sequence modeling on the language. King et al. (2020) examine sequence-to-sequence models and error analysis: "while these outperform traditional systems based on edit rule induction, it is hard to interpret what they are learning in linguistic terms" (p. 402). They developed a case study for Russian due to its low performance in SIGMORPHON 2016 and uncovered a "large class of errors in which the model incorrectly selects among lexically- or semantically-conditioned allomorphs" (King et al., 2020, p. 402). A robust error analysis allows the researcher to learn how a model differs from human behavior in its choices given that sequence-to-sequence models do not know what the inflected words *mean*. Perhaps most importantly, “our error analysis allows us to interpret what neural models are learning, reconnecting inflection tasks to linguistic intuitions by generalizing over error classes” (King et al., 2020, p. 404). One class of error noted that in certain classes of nouns “the accusative exhibits syncretism with either the genitive (for animates) or the nominative (for inanimates)” (King et al., 2020, p. 404). That is to say that the inflection of these nouns depends on semantic properties. In their model they concatenated word embeddings as a source of semantic information. Using the word embeddings almost halved the error rate of e-conjugation verbs, those verbs that take an —*e* for their second person singular form (King et al., 2020). Their emphasis on error-analysis and grouping of error classes inspires work that seeks to thoroughly interpret errors in sequence-to-sequence models.

2. Materials and Methods

This work continues the inquiry into Russian stress using modern machine-learning tools. We begin by first thoroughly describing the data and then describing the methods.

2.1 Description of the Data

The data used in the 8 experiments and sub-experiments was programmatically scraped from Giella project data (Moshagen et al., 2013) that was based on Zaliznyak's dictionary. Spektor (2021) created a lexicon called RusLex where the Giella infrastructure data was one of the four sources[1] for work in analyzing and detecting novel Russian loanwords. The data contain over 300,000 words, their morphological and part of speech features, the word's lemma, and the word with and without the lexical stress marking. Once collected, only symbols that were present in the dataset were removed as they did not contain stress and weren't pertinent to the experiments.

By part of speech, 47% were verbs, 32% were adjectives, 17% were nouns, 3% were determiners, 1% were pronouns, and the remainder were various types such as adverbs and interjections. 14% of all words were comparative adjectives, e.g. понорма́льней 'more normal', понебезопа́сней 'more secure', with the по- prefix.

From the dataset, the lemmas represent the dictionary form of a word before any kind of inflection. For verbs, which may be conjugated differently due to grammatical case, the lemma is shared in the dataset. For nouns, declension will lead to different forms for the singular and plural as well as the six grammatical cases, e.g. nominative, accusative, genitive, propositional, dative, and instrumental. These declined nouns will share a lemma, or dictionary form. As the

[1] https://github.com/undrits/ruslex or https://web.archive.org/web/20220510151619/https://github.com/undrits/ruslex/tree/main/ruslex

data takes into account morphology, many words, particularly verbs, have numerous entries due to unique inflection and gender, animacy, number, and case, e.g. one entry for the word штукова́вшей ‘concocted’ has the morphological tags V;Impf;PstAct;Fem;AnIn;Sg;Gen, while another entry for the same word has V;Impf;PstAct;Fem;AnIn;Sg;Loc. Two or more entries for a word may exist with the same morphology if their stress placement differs, e.g. антиле́нинская ‘anti-Leninist’ and а̀нтиле́нинская share the morphological tags A;Fem;AnIn;Sg;Nom. The words in the dataset span from the very familiar to the very technical. The word эшелони́ровавшие ‘to arrange in echelon formation’ has 24 entries, and the participle шосси́ровавшей ‘having traveled’ (among other meanings) has 48 entries. The imperfect verb шептать ‘to whisper’ has 708 distinct morphological entries (.23% of all words in the dictionary); conjugations include, e.g. шептал ‘whispered’, шепчущему ‘whispering’. The imperfect verb орать ‘to yell’ has 2,010 distinct morphological entries.

A few more examples of technical or alien words are: антисейсми́ческой ‘anti-seismic’, яровизировавшему ‘vernalizing: to cool a seed during germination’, побиолокацио́нней ‘more biolocational’, эмпириомони́змом ‘Empiriomonism: the philosophical work of Alexander Bogdanov from 1906’. Interestingly, as an example above, антиле́нинская ‘anti-Leninist’ is in the dictionary whereas ‘Leninist’ is not. Similarly, антиамерика́нск ‘anti-American’ is in the dictionary, but ‘American’ is omitted.

31,332 words (10% of all words) contain one or more secondary grave stress markers.

12,790 words (4% of all words) contain «ё». Stress is marked on «ё» although there are 2000 exceptions (16% of those containing «ё») in the data, e.g. шёлкотка́ной, or with secondary stress, потрё̀хме́сячней.

No primary stress was indicated in 1073 words (.35% of the total).

2.2 Methods

Derivation of Stress Codes

A stress code was derived for each word in the dataset. If the stressed vowel was at the end of the word a stress code of 0 signifying oxytone stress was assigned. Next, counting from the end of the word, the penultimate stress was given a 1, meaning a stress on the paroxytone. Next, if the antepenultimate syllable contained the stress, the word was assigned a 2, meaning a stress on the proparoxytone. The script continued until a stress code was assigned with the following exceptions: a NULL is assigned for those words without explicit stress markers; as discussed above, the «ё» is given primary stress, and the script halts and assigns a stress code when it discovers an «ё». Figure 1 shows the generation of the stress codes.

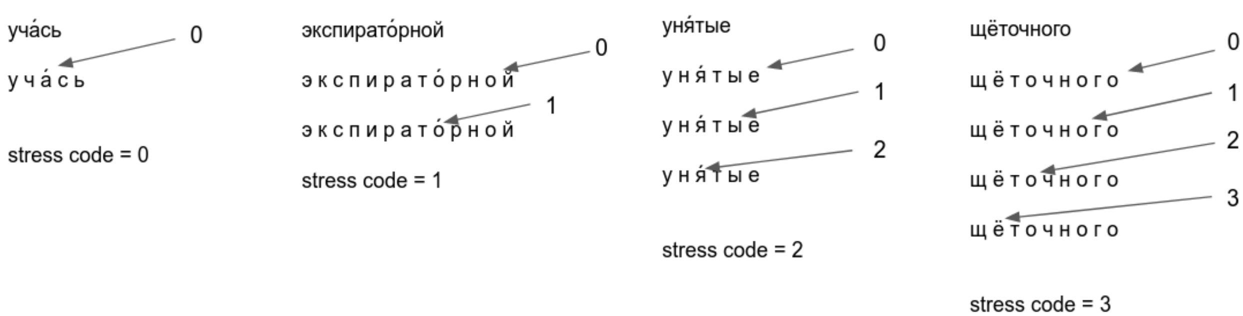

Figure 1 - Derivation of Stress Codes

Simply predicting paroxytone through the preantepenult is correct for 81% of the data.

To recall above, theorists had conflicting notions about where stress was likeliest to be located in adjectives. From the data: adjectives have an average length of 13 characters, suggesting that even with consonant clusters the majority of adjectives have the stress towards the end of the word or nearer the middle, e.g. понорма́льней ‘more normal’. Statistically: 33% of adjectives are paroxytone and 43% are proparoxytone.

Table 2 - Stress Code Data

Oxytone Stress	17,111	6%
Paroxytone Stress	75,196	25%
Proparoxytone Stress	101,209	34%
Fourth syllable from the end	68,878	23%
Fifth syllable from the end	31,325	10%
Sixth syllable from the end	6,378	2%
Seven or more syllables from the end	302	.1%

Table 2 shows the percentages for each stress placement from the data. Once the stress codes were generated, the primary data consists of a tab-separated values (TSV) file in which each row contains the word without stress markers, the word with stress markers, the derived stress code, the lemma, and all morphological features.

FairSeq Experiments

There were 8 main experiments using FairSeq, a sequence-modeling toolkit developed by Ott et al. (2019). The following table describes the experiments:

Table 3: Experiment Descriptions

The example used throughout is ахнувшим 'gasped' with the lemma ахнуть 'to gasp'
Experiment 1 - Baseline Prediction - Given the grapheme without a stress, predict stress placement Source: а х н у в ш и м Target: á х н у в ш и м
Experiment 2 - Stress Prediction with Stress Code - Given the grapheme with a trailing stress code, predict stress placement Source: а х н у в ш и м 2 Target: á х н у в ш и м
Experiment 3 - Stress Code Prediction - Given the grapheme, predict the stress code Source: а х н у в ш и м Target: 2
Experiment 4 - Stress Prediction with Lemma Feature α- Given the word's lemma, predict the stress placement Source: а х н у в ш и м ахнуть Target: á х н у в ш и м
Experiment 5 - Stress Prediction with Lemma Feature β- Given the word's lemma, predict the stress code Source: а х н у в ш и м ахнуть Target: 2
Experiment 6 - Stress Prediction with Morphological Features - Given all of the word's morphological features, predict the stress placement Source: а х н у в ш и м V Perf PstAct Neu AnIn Sg Ins Target: á х н у в ш и м
Experiment 7 - Stress Prediction with Morphological Feature (Part of Speech) - Given the word's part of speech feature, predict the stress placement Source: а х н у в ш и м V Target: á х н у в ш и м
Experiment 8 - Stress Code Prediction with Morphological Feature (Part of Speech) - Given the word's part of speech feature, predict the stress code Source: а х н у в ш и м V Target: 2

For each experiment conducted, the data was randomly shuffled and split into 80% training, 10% development/validation, and 10% test sets.

After preliminary adjustments to FairSeq parameters, each experiment had unique random seeds, a bidirectional encoder, an LSTM architecture, a maximum update of 800, and a batch-size of 3,000. The model was built using the Adam optimizer developed by Kingma & Ba (2014), a learning rate of .001, one encoder layer, and one decoder layer. Once trained, the neural network employed the model on the development/validation set and then the test set. The word error rate[2] (WER) was computed for each experiment. In order to be correct, each grapheme or stress code needs to be precisely accurate; a misplaced stress will result in that particular incorrect hypothesis raising the overall WER.

In addition to the experiments above, noting that the 10% of words with secondary stress led to a higher WER, secondary stress was removed from all experiments that had word-form targets and 5 experiments were then conducted in each; that is to say, if the target was a stress code, the experiment was not conducted again with the secondary stress removed. This is similar to the methods of Hall & Sproat (2013), who chose two conditions: to "ignore secondary stress in training and evaluation" and another experiment where they include secondary stress. As noted by Gorman & Bedrick (2019), evaluation with a standard split is insufficient for system comparison, and they propose "an alternative based on multiple random splits" (p. 2786). These further sub-experiments were conducted five times with unique random seeds on shuffled, randomly split data each time. The results are the median of the five experiments on the test set, noting the minimum and maximum results.

---

[2] The word error rate was computed using a script authored by Kyle Gorman that parses the output of *fairseq-generate*.

Some experiments below were designed to find the stress code as the target and some were designed to predict stress locations in the word as the target. While the stress code target performs assuredly well, the grapheme targets provide us insight into where the model falls short and specifically why. This error analysis with grapheme errors complements the stress code experiments. The mixed experiments therefore give us a more thorough understanding of stress-prediction errors and challenges for the model.

3. Results

Table 4: Results

Experiment Number and Brief Description	**Word Error Rate**
Experiment 1 - Baseline Prediction Source: а х н у в ш и м Target: á х н у в ш и м	22.28
Experiment 2 - Stress Prediction with Stress Code Source: а х н у в ш и м 2 Target: á х н у в ш и м	11.61
Experiment 3 - Stress Code Prediction Source: а х н у в ш и м Target: 2	13.93
Experiment 4 - Stress Prediction with Lemma Feature α Source: а х н у в ш и м ахнуть Target: á х н у в ш и м	16.37
Experiment 5 - Stress Prediction with Lemma Feature β Source: а х н у в ш и м ахнуть Target: 2	6.39
Experiment 6 - Stress Prediction with Morphological Features Source: а х н у в ш и м V Perf PstAct Neu AnIn Sg Ins Target: á х н у в ш и м	24.66
Experiment 7 - Stress Prediction with Morphological Feature (Part of Speech) Source: а х н у в ш и м V Target: á х н у в ш и м	22.53
Experiment 8 - Stress Code Prediction with Morphological Feature (Part of Speech) Source: а х н у в ш и м V Target: 2	16.14

Table 5: Results: Sub-experiments with all Secondary Stress Removed

Experiment Number and Brief Description	WER (Minimum)	WER (Maximum)	WER (Median)
Experiment 1 - Baseline Prediction Source: а х н у в ш и м Target: á х н у в ш и м	13.04	15.07	14.10
Experiment 2 - Stress Code Prediction α Source: а х н у в ш и м 2 Target: á х н у в ш и м	.78	1.09	.93
Experiment 4 - Stress Prediction with Lemma Feature α Source: а х н у в ш и м ахнуть Target: á х н у в ш и м	5.96	7.29	6.26
Experiment 6 - Stress Prediction with Morphological Features Source: а х н у в ш и м V Perf PstAct Neu AnIn Sg Ins Target: á х н у в ш и м	14.24	15.83	15.15
Experiment 7 - Stress Prediction with Morphological Feature (Part of Speech) Source: а х н у в ш и м V Target: á х н у в ш и м	13.48	15.34	14.86

4. Discussion

To best understand the results, we’ll examine the WER for each experiment as well as an error analysis from FairSeq’s generated output on the test set.

Experiment 1

Baseline Prediction
Source: а х н у в ш и м
Target: а́ х н у в ш и м

Table 6: Experiment 1

Target	Hypothesis	Notes
Correct		
т ё р ш и е	т ё р ш и е	Correctly omitted a stress marker due to the «ё»
о́ с т р ы е	о́ с т р ы е	Correctly placed stress on the onset
ю р о́ д с т в о в а в ш и м и	ю р о́ д с т в о в а в ш и м и	With many possible vowels to place the stress, correctly predicted the first «о»
Incorrect		
я с н е́ я	я́ с н е я	Incorrectly predicted stress on the onset
ш т ы р ё м	ш т ы́ р ё м	While the stress code script would halt at «ё», this hypothesis erroneously predicted a stress on «ы»
ш п а н о́ ю	ш п а́ н о ю	Incorrectly predicted proparoxytone, while correct stress is paroxytone
о̀ б л о н о́	о́ б л о н о	Predicted primary stress on the onset, but the correct stress was the oxytone; leads to further errors with the secondary stress

Table 7: Experiment 1 with Secondary Stress Removed

Target	Hypothesis	Notes
Correct		
э к з о т и́ з м	э к з о т и́ з м	Correctly predicted that -изм always receives stress on the «и» as noted by Ward (1965, pp 48-49).
э к о л о́ г и и	э к о л о́ г и и	Correctly predicted that the declined suffix -ология (-ology) receives stress on the second «о» (Ward 1965, pp 48-49)
э л е к т р о д о́ й к а х	э л е к т р о д о́ й к а х	Correctly predicted paroxytone stress
Incorrect		
э г о ф у т у р и́ з м	э г о ф у т у́ р и з м	Incorrectly placed stress on the paroxytone, while the «и» in the suffix -изм should always receive the stress (Ward 1965, pp 48-49)
ш и́ т о г о	ш и т о́ г о	Incorrectly predicted paroxytone, but proparoxytone was correct
ш е л к а́ х	ш е́ л к а х	Incorrectly predicted paroxytone, but oxytone was correct

The baseline experiment shows how difficult and irregular the Russian language is for correct stress placement. All parts of speech were present, and no features were provided. The model performed with a WER of 22.28 in the first experiment, meaning that it was incorrect in precise prediction nearly a quarter of the time. Once secondary stress markers were removed, the WER dropped to 14.10.

Experiment 1 provides insight into some of the errors in correctly predicting the placement of stress. With no further input, it's difficult to train a model that will provide precise hypotheses.

The suffix -изм was correctly predicted in 11 of 12 instances in the test set.

Keeping this baseline in mind, we move on to stress codes and features.

Experiment 2

Stress Prediction with Stress Code
Source: а х н у в ш и м 2
Target: а́ х н у в ш и м

Table 8: Experiment 2

Target	Hypothesis	Notes
Correct		
о т п е р л с я́	о т п е р л с я́	Correctly predicted the oxytone stress
ё ж и в ш е м с я	ё ж и в ш е м с я	With a stress code of 3, the model correctly predicted the «ё» as the primary stress
п о́ п к а м	п о́ п к а м	Correctly predicted a paroxytone stress
Incorrect		
ё̀ ж - р ы́ б у	ё ж - р ы́ б у	The stress code system is unaware of placing secondary stress properly; the stress code script halted at the paroxytone «ы»
а̀ г р о х и́ м и к	а г р о х и́ м и к	Secondary stress not placed properly
а в т о ш к о́ л ы	а̀ в т о ш к о́ л ы	Incorrectly hypothesized a secondary stress on the onset

Table 9: Experiment 2 with Secondary Stress Removed

Target	Hypothesis	Notes
Correct		
п о т р ё х д н é в н е е	п о т р ё х д н é в н е е	Correctly predicted the stress on the proparoxytone but the script didn't reach the «ё»
у н т á х	у н т á х	Correctly predicted stress on the oxytone
а н г л о с á к с а	а н г л о с á к с а	Correctly predicted paroxytonal stress
Incorrect		
а г р о л е с о м е л и о р а т и́ в н о ю	а г р о л е с о м е л и о р а т и в н о ю	With a stress code of 2, the proparoxytone should have been marked, but the hypothesis was for no stress markings
п о т р ё х á к т н е е	п о т р ё х а к т н е е	With a stress code of 2, the model did not place stress on the proparoxytone due to the presence of the «ё»
э с э́ с о в с к о г о	э́ с с о в с к о г о	The model incorrectly hypothesized deleting the «э» and placing stress on the onset

Experiment 2 was designed to test the efficacy of the stress code derivation. We predicted that given the stress code as a feature, the neural network should predict placement of the stress almost perfectly. If it did not, then perhaps the derivation method could be improved. While a Python script could easily perform this task, processing this through the LSTM pointed to ambiguities within the dataset and our assumptions about default stress on «ё».

The experiment had a WER of 11.61 and a WER of .93 with secondary stress removed.

The model improved greatly once secondary stress was removed. The errors that remain show inconsistencies between the script that derives the stress code and the data from which it trained. In words such as ё̀ж-ры́бу 'hedgehog fish', the model wouldn't know to place a grave accent on the already-stressed «ё». With all secondary stress removed, there are still similar errors involving «ё» that account for why, although a very good predictor with a WER of .93, it is not perfect in its hypotheses.

Experiment 3

Stress Code Prediction
Source: а х н у в ш и м
Target: 2

Table 10: Experiment 3

Grapheme	Target Stress Code	Hypothesis	Notes
		Correct	
о п у с а х	2	2	Correctly predicted stress on the onset
ш и т ь я м	0	0	Correctly predicted oxytone stress
щ е н к а м	NULL	1	0.35% of the words contain no stress data
		Incorrect	
и в у ш к у	2	1	Incorrectly predicted stress on the paroxytone
ш а ф е р а	0	2	Incorrectly predicted the onset, while stress was on the oxytone
ш т р и х к о д о в	1	0	Incorrectly predicted oxytone stress

This experiment was designed to determine the stress code given only the unstressed word. To recall: simply predicting paroxytone (stress code 1) through the preantepenult (stress code 3) is correct for 81% of the data.

With a WER of 13.93, this experiment exhibits the challenges to modeling stress in Russian to a stress code only. There are no features present from which to train.

Experiment 4

Stress Prediction with Lemma Feature α
Source: а х н у в ш и м ахнуть
Target: á х н у в ш и м

Table 11: Experiment 4

Target	Hypothesis	Lemma	Notes
Correct			
ш и б á т ь	ш и б á т ь	шибать	The lemma is the same as the grapheme, and the stress was correctly predicted
ó р у щ и м	ó р у щ и м	орать	Correctly predicted proparoxytone stress
о р ý щ е й	о р ý щ е й	орать	Correctly predicted paroxytone stress - the lemma орать occurs 179 times in the test set
à н г л о я з ы́ ч н о	à н г л о я з ы́ ч н о	англоязычный	Correctly predicted primary and secondary stress
à р т у ч и́ л и щ а м	à р т у ч и́ л и щ а м	артучилище	Correctly predicted primary and secondary stress
à в т о с т о я́ н к и	à в т о с т о я́ н к и	автостоянка	Correctly predicted primary and secondary stress
à в т о в а к ц и́ н у	à в т о в а к ц и́ н у	автовакцина	Correctly predicted primary and secondary stress
н и п у т ё̀ м ч е г ó	н и п у т ё м ч е г ó	ничто	Correctly predicted oxytone, but did not give secondary stress to the «ё»

ш ё̀ л к о т к а́ н о ю	ш ё л к о т к а́ н о ю	шёлкотканый	Correctly predicted proparoxytone, but did not give secondary stress to the «ё»
ш и п у н а́ м и	ш и п у н а́ м и	шипун	Correctly predicted the paroxytone stress for the suffix -ами
Incorrect			
о́ р у щ и х	о р у́ щ и х	орать	Incorrectly predicted paroxytone stress
у́ ч а щ и х	у ч а́ щ и х	учить	Incorrectly predicted paroxytone stress
ш и́ п а м и	ш и п а́ м и	шип	Incorrectly predicted paroxytone stress

Table 12: Experiment 4 with secondary stress removed

Target	Hypothesis	Lemma	Notes
Correct			
я́ х т н о м	я́ х т н о м	яхтный	Correctly predicted paroxytone stress
О́ л ь г а х	О́ л ь г а х	Ольга	Correctly predicted the paroxytone stress
ё ж и к а м	ё ж и к а м	ёжик	Correctly predicted primary stress on «ё»
Incorrect			
о́ б н я т ы м и	о б н я́ т ы м и	обнять	Incorrectly predicted a proparoxytone stress
о́ т н я т о й	о т н я́ т о й	отнять	Incorrectly predicted paroxytone stress
а́ л ь т а м и	а л ь т а́ м и	альт	Incorrectly predicted paroxytone stress

Certainly training on lemma data improved the model. The model can identify prefixes, suffixes, and conjugations by being provided the lemma as a feature. By knowing that the lemma for шелепом is шелеп ‘whisper’, the model was able to learn that -ом was the suffix. On the other hand in a different example, the prefix авто- ‘auto-, self-’ is given secondary stress on the onset, and the model predicted that correctly despite the source data containing both forms with identical morphologies: e.g. автотра́нспортное and а̀втотра́нспортное ‘auto transport’. The lemma included the prefix so there was no way for the model to drill down into the word and identify, e.g. авто- ‘auto-’ or анти- ‘anti-’ as a prefix.

Two words that shared the lemma орать ‘to yell’ differed only in the verb’s conjugation. The stress was correctly predicted for the word ору́щей, but predicted incorrectly when the mobile stress moved to the onset in óрущих. We may also recall that there are 2,010 distinct morphological entries for this lemma in the source data.

For the word áльтами ‘violas’, the model predicted paroxytone stress; the suffix -ами is the Russian instrumental case for plurals following a consonant in the nominative singular form. The dataset contains a homograph of this word with the stress on the paroxytone: альтáми ‘altos’[3]. Clearly a model will find it difficult to place stress correctly when trained on homographs with different marked stress and the shared lemma альт. The ambiguity of this stress could point to possible causes: the stress is changing or has changed since Zaliznyak’s dictionary in 1977; the word could simply be infrequent or unfamiliar. Since the dataset does not include definitions, it’s difficult to be certain which word means which definition.

The model went from a WER of 16.37 to 6.26 with secondary stress removed.

[3] When I asked three native Russian speakers about this example, one said stress on the first vowel for ‘viola’ and second for ‘alto’; one said second vowel in both cases; the third said first vowel in both cases.

Experiment 5

Stress Prediction with Lemma Feature β
Source: а х н у в ш и м ахнуть
Target: 2

Table 13: Experiment 5

Source	Hypothesis	Notes
Correct		
щ ё л к а ю щ и е с я щёлкать	5	Correctly predicted the «ё» present in the lemma
щ е р и в ш е г о с я щериться	4	Correctly predicted stress on the onset
ш о р к а х шорка	1	Correctly predicted oxytone stress
Incorrect		
а н т и м и р о м антимир	0	Incorrectly predicted oxytone stress, while correct stress is on the paroxytone
ш л ю п о к шлюпка	0	Incorrectly predicted oxytone stress while the correct stress is on the paroxytone - in two other instances with this lemma, the model correctly predicted paroxytone
и к о с о в икос	1	Incorrectly hypothesized paroxytone, but proparoxytone/onset was correct

This experiment shows that by providing an unstressed word with its lemma as a feature, we get a WER of 6.39.

Experiment 6

Stress Prediction with Morphological Features
Source: а х н у в ш и м V Perf PstAct Neu AnIn Sg Ins
Target: а́ х н у в ш и м

Table 14: Experiment 6

Target	Hypothesis	Morphology	Notes
Correct			
о́ б щ е й	о́ б щ е й	A Fem AnIn Sg Loc	Correctly predicted paroxytone stress
ш к о д л и́ в е е	ш к о д л и́ в е е	A Cmpar Pred	Correctly predicted comparative adjective with proparoxytone stress
ш ѐ р с т о ч е с а́ л ь н а я	ш е р с т о ч е с а́ л ь н а я	A Fem AnIn Sg Nom	Correctly predicted proparoxytone stress but failed to predict placement of secondary stress
о́ б у х у	о́ б у х у	N Msc Inan Sg Dat	Correctly predicted proparoxytone stress
ш о́ р к и	ш о́ р к и	N Fem Anim Pl Nom	Correctly predicted paroxytone stress
ш и́ л а	ш и́ л а	V Impf Pst Fem Sg	Correctly predicted paroxytone stress
и́ з б р а н н о й	и́ з б р а н н о й	V Perf TV PstPss Fem AnIn Sg Ins	Correctly predicted stress on the onset
и́ з б р а н н о й	и́ з б р а н н о й	V Perf TV PstPss Fem AnIn Sg Ins	Correctly predicted stress on the onset
Incorrect			
п о т р ё х с т в о́ р ч а т е е	п о т р ё х с т в о р ч а́ т е е	A Cmpar Pred Att	Incorrectly predicted proparoxytone stress - added stress despite presence of «ё»

ш р ѝ- л а н к и́ й с к о г о	ш р и - л а н к и́ й с к о г о	A Msc AnIn Sg Gen	The prefix шри- for Sri-Lankan has entries for both шрѝ- and шри in the dataset
а̀ н т и ч а с т и́ ц е ю	а н т и ч а с т и́ ц е ю	N Fem Inan Sg Ins Fac	Correctly predicted primary stress placement but failed to place secondary stress for the prefix анти-
а н т и л е́ н и н с к и м	а н т и л е н и́ н с к и м	A MFN AnIn Pl Dat	Incorrectly predicted paroxytone stress
п о а л е́ е	п о а́ л е е	A Cmpar Pred Att	Incorrectly predicted comparative adjective
а в т о д о е́ н и я	а в т о д о́ е н и я	N Neu Inan Sg Gen	Incorrectly predicted proparoxytone stress
о́ т п е р ш е й с я	о т п е́ р ш е й с я	V Perf IV Der Der/PstAct A Fem AnIn Sg Ins	Incorrectly predicted stress on the first «е» instead of on the onset
у́ з и в ш и х	у з и́ в ш и х	V Impf IV Der Der/PstAct A MFN Anim Pl Acc	Incorrectly predicted paroxytone stress instead of proparoxytone/onset

Table 15: Experiment 6 with secondary stress removed

Target	Hypothesis	Morphology	Notes
Correct			
э л и́ т н а	э л и́ т н а	A Fem Sg Pred	Correctly predicted paroxytone stress
п о г у м м о́ з н е й	п о г у м м о́ з н е й	A Cmpar Pred Elid Att	Correctly predicted paroxytone stress
А́ й б о л и т а м	А́ й б о л и т а м	N Prop Msc Anim Pl Dat	Correctly predicted stress on the onset
а г р о б о т а́ н и к о ю	а г р о б о т а́ н и к о ю	N Fem Inan Sg Ins Fac	Correctly predicted stress on the «а»
у́ с л а н н у ю	у́ с л а н н у ю	V Perf PstPss Fem AnIn Sg Acc	Correctly predicted stress on the onset
ш р е д е р у́ е м	ш р е д е р у́ е м	V Impf PrsPss Msc Sg Pred	Correctly predicted paroxytone stress
Incorrect			
щ а н ы́ м	щ а́ н ы м	A Msc AnIn Sg Ins	Incorrectly predicted paroxytone stress
я м щ и́ ч ь е й	я́ м щ и ч ь е й	A Fem AnIn Sg Loc	Incorrectly predicted onset stress
я т о́ в и	я́ т о в и	N Fem Inan Pl Acc	Incorrectly predicted onset
ш а с л а́	ш а́ с л а	N Fem Inan Sg Nom	Incorrectly predicted paroxytone stress instead of oxytone
ш и п я́ щ и м	ш и́ п я щ и м	V Impf IV PrsAct MFN AnIn Pl Dat	Incorrectly predicted proparoxytone stress
ш и п я́ щ е ю	ш и́ п я щ е ю	V Impf PrsAct Fem AnIn Sg Ins Leng	Incorrectly predicted onset instead of proparoxytone stress

Experiment 6 was designed to determine if training the model with the complete morphology as a feature would improve prediction results.

The model with retained secondary stress had a WER of 24.66 which is slightly worse than Experiment 1, our baseline experiment with no features at all. With secondary stress removed, the model had a WER of 15.15, significantly better but still worse than the WER of 14.10 from Experiment 1 with secondary stress removed. Perhaps overfitting has occurred, establishing complexity of the model that impares correct prediction.

Experiment 7

Stress Prediction with Morphological Feature (Part of Speech)
Source: а х н у в ш и м V
Target: а́ х н у в ш и м

Table 16: Experiment 7

Target	Hypothesis	Part of Speech	Notes
Correct			
я́ ч н о г о	я́ ч н о г о	Adjective	Correctly predicted onset stress
э с к у́ д о	э с к у́ д о	Noun	Correctly predicted paroxytone stress
и́ х н е м у	и́ х н е м у	Pronoun	Correctly predicted onset stress
щ и́ п а н о	щ и́ п а н о	Verb	Correctly predicted onset stress
Incorrect			
я м с к о́ г о	я́ м с к о г о	Adjective	Incorrectly predicted stress on onset, but stress is on the -ого

			(genitive case, masc. or neut.) suffix
ю́ т о в о г о	ю т о в о́ г о	Adjective	Incorrectly predicted paroxytone, but stress is on the onset
о м у л е́ й	о м у́ л е й	Noun	Incorrectly predicted paroxytone stress instead of oxytone
щ а в е́ л ь н и к	щ а́ в е л ь н и к	Noun	Incorrectly predicted onset instead of paroxytone
э́ д а к о г о	э д а к о г о́	Pronoun	Incorrectly predicted oxytone instead of onset
э́ т а к о г о	э т а к о́ г о	Pronoun	Incorrectly predicted paroxytone instead of onset
о́ т н я т ы	о т н я́ т ы	Verb	Incorrectly predicted paroxytone stress instead of onset.
э́ к а в ш е й	э к а́ в ш е й	Verb	Incorrectly predicted paroxytone instead of onset
ш о ф е р я́ щ и х	ш о ф е́ р я щ и х	Verb	Incorrectly predicted proparoxytone instead of paroxytone

Table 17: Experiment 7 with secondary stress removed

Target	Hypothesis	Part of Speech	Notes
Correct			
у́ с т н о е	у́ с т н о е	Adjective	Correctly predicted proparoxytone stress
ш у́ ш е р у	ш у́ ш е р у	Noun	Correctly predicted proparoxytone stress
и́ х н и й	и́ х н и й	Pronoun	Correctly predicted paroxytone stres
щ у́ р я щ и й с я	щ у́ р я щ и й с я	Verb	Correctly predicted stress on the onset
Incorrect			
я м с к о́ й	я́ м с к о й	Adjective	Incorrectly predicted onset stress instead of oxytone stress
ш и́ б к о е	ш и б к о́ е	Adjective	Incorrectly predicted oxytone instead of onset
п о к а л а н ч о́ в е й	п о к а л а́ н ч о в е й	Adjective	Incorrectly predicted a stress on the «а» while stress is nearer the end on the proparoxytone
о б л о н о́	о б л о́ н о	Noun	Incorrectly predicted paroxytone instead of oxytone
ш а р а́ м и	ш а́ р а м и	Noun	Incorrectly predicted proparoxytone stress instead of paroxytone stress on the suffix -ами (plural, instrumental case)
я́ р к а м и	я р к а́ м и	Noun	Incorrectly predicted paroxytone stress

			instead of onset
н е́ н а к о м	н е н а к о́ м	Pronoun	Incorrectly predicted oxytone instead of onset
э́ д а к о м у	э д а к о́ м у	Pronoun	Incorrectly predicted paroxytone instead of onset
о́ б н я т о м	о б н я́ т о м	Verb	Incorrectly predicted paroxytone instead of onset
с е к л и́ с ь	с е́ к л и с ь	Verb	Incorrectly predicted onset instead of oxytone
и́ з б р а н а	и з б р а́ н а	Verb	Incorrectly predicted paroxytone instead of onset

In an effort to minimize overfitting due to too many features, this experiment was designed to take one part of speech feature to predict correct stress placement.

For this experiment with secondary stress intact, the WER was 22.53. From the examples in Table 16, the mobile stress was not predicted well for adjectives with the -oro suffix for masculine and neuter in the genitive case. 845 occurrences of adjectives with the -oro suffix were present in the test set.

In the 5 experiments with secondary stress removed, the WER median was 14.86 (minimum of 13.48, maximum of 15.34), an improvement, but suffering from the same problems with predicting mobile stress. Experiment 7 performed only slightly better than Experiment 1 with no features and a WER of 22.28. With secondary stress removed, Experiment 7 had a WER of 14.86 and performed worse than Experiment 1 with a WER of 14.10 with secondary stress removed and no features.

Experiment 8

Stress Code Prediction with Morphological Feature (Part of Speech)
Source: а х н у в ш и м V
Target: 2

Table 18: Experiment 8

Source	Hypothesis	Notes
Correct		
э п и ч е с к а A	2	Correctly predicted proparoxytone stress
щ ё л о ч и N	2	Correctly predicted proparoxytone stress
щ ё л к а е м ы й V	3	Correctly predicted stress on the onset «ё»
Incorrect		
э л ь ф о в N	0	Incorrectly predicted oxytone instead of paroxytone
щ е п н е е A	2	Incorrectly predicted proparoxytone instead of paroxytone stress
э л и с т и н с к и A	2	Incorrectly predicted proparoxytone stress instead of paroxytone
ш у т к о в а т ь с я V	2	Incorrectly predicted proparoxytone stress instead of paroxytone stress
ш у с т р и в ш и м и V	3	Incorrectly predicted onset stress instead of proparoxytone stress

This experiment was designed to use the part of speech feature to predict a stress code. The WER was 16.14. This experiment is similar to the output of experiment 3 that had a WER of 13.93; adding the part of speech feature led to worse results when predicting a stress code.

5. Threats to Validity

The data was not split using lexeme-aware methods as described by Gorman et al. (2020): "the splitting procedure was constrained so that all inflectional variants of any given lemma [...] are limited to a single shard" (p. 42). As mentioned above, шептать 'to whisper' with its 708 derivations in the data could very well exist in training as well as development and test sets; this could skew the results as the model had seen the data before. Lexeme-aware splitting would assure us that words sharing lemma are placed in the same set. This kind of splitting is important especially for experiments that use lemma as a feature, and the other experiments indirectly. Of course, mobile stress occurs on words with shared lemma, so having trained on shared lemma does not guarantee correct stress-placement. While this is a small threat to validity, given the size of the data, the findings above are still quite significant.

To recall, no stress was indicated in 1073 words (.35% of the total). This data was not removed from the set. The words likely contributed to higher word error rates as the model was largely trained to place a primary stress marker.

The LSTM model was unaware of any secondary stress related to hyphenated words. Hyphenated compounds tend to have double-stress. The secondary stress is always on the first part of the hyphenated compound and the primary stress on the second part. As the stress codes were derived from the end of the word and our goal was to place primary stress, this did not affect the results of the experiments with secondary stress removed.

6. Future Work and Conclusion

Whether the target was the placement of primary stress or predicting a stress code, our LSTM neural network produced the best WER when given the lemma as a feature. Experiment 4 with secondary stress removed resulted in a WER of 6.26. Experiment 5, which predicted a stress code, had a WER of 6.39. These were the two best results in the experiments.

Stress homographs, particularly those with secondary stress markers, accounting for more than 10% of all words, played a significant role in raising the word error rate in all experiments when retained. For example from the dataset, the adjective with the lemma общезаводской 'common for the whole factory; plant-wide' may be pronounced in four unique ways: общезаводско́ю, общезаво́дскою, о̀бщезаводско́ю, or о̀бщезаво́дскою. Marks of secondary stress are much more prominent than stress homographs on the primary stress. A noun such as шкиву́ 'pulley' may also be pronounced шки́ву with identical morphology and lemma. This makes it difficult for machine-learning to predict stress placement using solely this dataset.

If we compare our results to paradigmatic work described above, we see that indeed Coats (1976) was correct: stem-stressed words are indeed difficult to predict even when the "phonological, morphological, syntactic, or semantic properties" (p. 2) are known. While above, Shapiro (1986) and others maintained that "membership in a particular form class or semantic category" (p. 183) narrows the possible accent positions, whether due to overfitting with the full morphology in Experiment 6 or due to membership not being a reliable sole predictor, neither full morphology as a feature nor part of speech as a feature provided exemplary results.

The work of Hart (2015), described above, focused on the -ик (masculine) and -ица (feminine) suffixes for nouns. To recall, Hart (2015) found that "stress in Russian derived nouns functions to define the boundary between the root and the suffix" (p. 1), and nouns with these suffixes have stress on the same vowel throughout the paradigm. By training on the lemma feature and having seen the lemma before with stress on a certain vowel, the model was able to learn that characteristic and place stress correctly.

King et al. (2020) raised some very interesting prospects for animacy and inanimacy as a feature for nouns, and surprisingly also for verbs. While Experiment 6 attempted to provide morphological features to aid in prediction, it had the worst results of the experiments. Future work will take animacy and inanimacy as sole features. Also, their work in grouping error classes would be most useful for a stress-assignment task as the one presented here. The work of Reynold & Tyers (2015) also provides inspiration for conceiving of a system of error-tolerance in which non-confident predictions could be abstained.

As our data is in essence a synchronic dictionary-sourced time-capsule of Russian, our neural network results are, too, synchronic. Hingley (1952) noted this as he mentioned hesitant stress, a transitional stress location. Lagerberg (2007) in his work on stress and frequency, pointed to the sociolinguistic work of Sharapova (2000) using surveys of native speakers to observe where stress is actually placed. Sharapova (2000) summed it up: "the compilers of stress handbooks do not have a defined norm concept, but see *norms* rather as an opportunity to have one's subjective pick from predecessors' recommendations" (p. 93). Jouravlev & Lupker (2015) in a similar vein stated "real language usage might deviate significantly from the canonical, prescribed usage reflected in the dictionaries" (p. 949).

As mentioned above and in the literature (Shapiro, 1986; Lagerberg, 2007), frequency is an important, but not the sole, factor in determining stress position. The Zaliznyak dataset contains no frequency data. Frequency data as a feature would be an improvement to work in this area. Further work may be done as well in updating Zaliznyak's dictionary with modern pronunciation and adding neologisms.

References

Arciuli, J., & Cupples, L. (2006). The processing of lexical stress during visual word recognition: Typicality effects and orthographic correlates. *The Quarterly Journal of Experimental Psychology, 59*(5), (pp. 920-948). https://doi.org/10.1080/02724980443000782

Coats, H. S. (1976). *Stress assignment in Russian*. Edmonton, Alta: Linguistic Research.

Derbyshire, W. W. (1967). Verbal homonymy in the Russian language. *Canadian Slavonic Papers*, *9*(1), (pp. 131-139).

Dou, Q., Bergsma, S., Jiampojamarn, S., & Kondrak, G. (2009). A ranking approach to stress prediction for letter-to-phoneme conversion. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP* (pp. 118-126).

Feldstein, R. (1980). On the definition of Russian stress paradigms. *General Linguistics*, *20*(3), (pp. 123-139).

Feldstein, R. (1986). The Russian verbal stress system. *International Journal of Slavic Linguistics and Poetics*, *XXXIII*, 43-61. https://pollang.sitehost.iu.edu/ijslp23.pdf

Feldstein, R. (1993). The nature and use of the accentual paradigm as applied to Russian. *Journal of Slavic Linguistics*, *1*(1), (pp. 44–60). http://www.jstor.org/stable/24599037

Feldstein, R. (1996). Regular and deviant patterns of Russian nominal stress and their relationship to markedness. In Andrews, E. & Tobin, Y. (Eds.), *Toward a calculus of meaning: Studies in markedness, distinctive features and deixis*. John Benjamins Publishing Company.

Gorman, K., Mazovetskiy, G., & Nikolaev, V. (2018). Improving homograph disambiguation with supervised machine learning. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Gorman, K., & Bedrick, S. (2019). We need to talk about standard splits. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 2786-2791).

Gorman, K., Ashby, L. F., Goyzueta, A., McCarthy, A. D., Wu, S., & You, D. (2020). The SIGMORPHON 2020 shared task on multilingual grapheme-to-phoneme conversion. In *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology* (pp. 40-50).

Hall, K., & Sproat, R. (2013). Russian stress prediction using maximum entropy ranking. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 879-883).

Halle, M. (1973). The accentuation of Russian words. *Language*, *49*(2), (pp. 312-348). https://doi.org/10.2307/412457

Haraldsson, H. (1996). Problems of dictionary grammar. The Zaliznyak solution: a boon or a burden? In *Proceedings of the 7th EURALEX International Congress* (pp. 561-572).

Hart, D. (2015). The stress of Russian nouns in -ик and -ица. *Poljarnyj Vestnik: Norwegian Journal of Slavic Studies*, 18 (October), (pp. 1-17). https://doi.org/10.7557/6.3451

Hingley, R. (1952). The stress of Russian nouns in-а/-я under inflection. *The Slavonic and East European Review*, *31*(76), (pp. 186-203).

Iosad, P., Koptjevskaja-Tamm, M., Piperski, A., & Sitchinava, D. (2018). Depth, brilliance, clarity: Andrey Anatolyevich Zaliznyak (1935–2017). *Linguistic typology*, *22*(1), (pp. 175-184).

Jakobson, R. (1948). Russian conjugation, WORD, 4:3, (pp. 155-167). https://doi.org/10.1080/00437956.1948.11659338

Jakobson, R. (1957). The relationship between genitive and plural in the declension of Russian nouns. *Scando-Slavica*, *3*(1), (pp. 181-186).

Jones, D. (1956). *An outline of English phonetics: 8th ed.,* Cambridge [Eng.] Heffer.

Jouravlev, O., & Lupker, S. J. (2014). Stress consistency and stress regularity effects in Russian. *Language, Cognition and Neuroscience*, *29*(5), (pp. 605-619).

Jouravlev, O., & Lupker, S. J. (2015). Predicting stress patterns in an unpredictable stress language: The use of non-lexical sources of evidence for stress assignment in Russian. *Journal of Cognitive Psychology*, *27*(8), (pp. 944-966). https://doi.org/10.1080/20445911.2015.1058267

Kempgen, S. (1990). Review of the book *Report on Russian morphology as it appears in Zaliznyak's grammatical dictionary* by E. Ilola and A. Mustajoki. *Russian Linguistics*, 14(2), (pp. 209-211).

King, D. L., Sims, A. D., & Elsner, M. (2020). Interpreting sequence-to-sequence models for Russian inflectional morphology. *Proceedings of the Society for Computation in Linguistics*, *3*(1), (pp. 402-411).

Kingma, D. and Ba, J. (2015) Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*. Preprint available via arXiv at https://arxiv.org/abs/1412.6980

Lagerberg, R. (2007). Variation and frequency in Russian word stress. *Australian Slavonic and East European Studies (ASEES) Journal*, 21(1-2), (pp. 165-176). https://miskinhill.com.au/journals/asees/21:1-2/variation-and-frequency-in-stress.pdf

Lavitskaya, Y., & Kabak, B. (2014). Phonological default in the lexical stress system of Russian: Evidence from noun declension. *Lingua*, *150*, (pp. 363-385).

Moshagen, Sjur N., Tommi Pirinen, and Trond Trosterud. (2013). Building an open-source development infrastructure for language technology projects. In *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013)*, (pp. 343–352).

Mustajoki, A. (1981). Russian stress: Word frequency as an indicator of mobile stress. *Scando-Slavica*, *27*(1), (pp. 199-211).

Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N,. Grangier, D., and Auli, M. 2019. fairseq: A fast, extensible toolkit for sequence modeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, Minneapolis, Minnesota. Association for Computational Linguistics, (pp. 48-53).

Ponomareva, M., Milintsevich, K., Chernyak, E., & Starostin, A. (2017). Automated word stress detection in Russian. In *Proceedings of the First Workshop on Subword and Character Level Models in NLP*, (pp. 31–35). https://doi.org/10.18653/v1/W17-4104

Protopapas, A., Gerakaki, S., & Alexandri, S. (2007). Sources of information for stress assignment in reading Greek. *Applied Psycholinguistics*, *28*(4), (pp. 695-720).

Reynolds, R., & Tyers, F. (2015). Automatic word stress annotation of Russian unrestricted text. In *Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015)*, (pp. 173-180).

Shapiro, M. (1986). The Russian system of stress. *Russian Linguistics*, *10*(2), (pp. 183-204). http://www.jstor.org/stable/40160135

Sharapova, E. M. (2000). *Implicit and explicit norm in contemporary Russian verbal stress* (Doctoral dissertation, Acta Universitatis Upsaliensis, Uppsala, Sweden). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-1206

Spektor, Y. (2021). *Detection and morphological analysis of novel Russian loanwords* (Master's thesis, CUNY Graduate Center, New York, NY). Retrieved from https://academicworks.cuny.edu/gc_etds/4572/

Wade, T., & Gillespie, D. (2011). *A comprehensive Russian grammar*. Wiley-Blackwell.

Ward, D. (1965). *The Russian language today: System and anomaly*. Chicago: University Press.

Zaliznyak, A.A. (1977). Grammatičeskij slovar' russkogo jazyka. Slovoizmenenie [A grammatical dictionary of Russian: Inflection]. Moscow: Russkij jazyk.