Computational Study of Shughni Phonotactics

Yury Makarov
HSE University / Moscow, 21/4
Staraya Basmannaya Ulitsa
Institute of Linguistics, RAS / Moscow,
Bolshoy Kislovsky Lane, 1/1
yurmak@iling-ran.ru

Abstract

Detailed descriptions of Shughni phonotactics are scarce. It does not come as a surprise: not only Shughni (an Iranian language spoken in the Pamir Mountains) has a relatively small number of speakers (ca. 100,000) but up until recently, no databases were facilitating phonological research on it. Now that *pamiri.online*, a website on Pamir languages, has been developed, new data can be used for studying the sound patterns of Shughni. This paper illustrates how *pamiri.online* can be employed to update and enhance the descriptions of Shughni phonotactics.

Keywords: Iranian languages; Shughni; phonology; phonotactics

Изучение шугнанской фонотактики цифровыми методами

Юрий Юрьевич Макаров
НИУ ВШЭ / Москва,
Старая Басманная, 21/4
ИЯз РАН / Москва,
Б. Кисловский пер., 1с1
yurmak@iling-ran.ru

Аннотация

Подробные описания шугнанской фонотактики практически отсутствуют. Это неудивительно: на шугнанском языке (иранская группа; основной ареал — Западный Памир) говорит сравнительно малое число человек (ок. 100000) и до недавнего времени не было цифровых баз данных, которые способствовали бы его фонологическим исследованиям. Платформа *pamiri.online*, посвященная памирским языкам в целом, сделала возможным применение автоматических методов анализа к шугнанскому материалу, и цель этой статьи — продемонстрировать некоторые результаты, полученные таким способом.

Ключевые слова: иранские языки; шугнанский язык; фонология; фонотактика

1 Literature review

There are only a few descriptions of Shughni phonotactics. In [1, p. 33–34], which concerns only syllable structure, it is claimed that only these seven syllables occur in Shughni (transcriptions and translations by Olson):

(1)	V	VC	CV	VCC	CVC	CVCC	CVCCC
	/ø/	/at/	/tʊ/	/arz/	/ʒiz/	/vɔrd͡ʒ/	/ʃarθk/
	‘hey’	‘and’	‘you’	‘suggestion’	‘wood’	‘horse’	‘mud’

Olson notes that all these syllables can stand alone as words, though V is not common. Then, she mentions that ‘VCC was not found in the first syllable of a word with more than one syllable’ (p. 33),

unlike all the other syllables. As for word-medial position, CV, CVC and CVCC are attested. Finally, CV, CVC, CVCC and CVCCC are found in final position in polysyllabic words.

Olson summarizes her findings by stating that 'Shughni allows complex syllable codas syllable-finally, though these codas are more rare non-word-finally… Onsets must be a single consonant (or single vowel word-initially). They must not be a consonant cluster' (p. 34).

In [2], there is only a short note on clusters:

> "In general, there are no initial clusters. Final clusters in loans which deviate from permitted indigenous clusters are assimilated by release vowels: *umr* > *u̲mri* 'life', *naql* > *na̲qli* 'narration'."

In [3], which concerns the Shughni-Rushani subgroup in general, CV, CVC and CVCC are mentioned as 'optimal'; V, VC, VCC, CV, CVC, CVCC, CVCCC are attested (coda clusters are said to be organized according to the principle of decreasing sonority and airflow, and increasing intensity of articulation). These aspects of Shughni-Rushani phonotactics, according to Edelman, led to some synchronic and diachronic sound changes (transcriptions and translations by Edelman):

1) Prothetic [h j w] before initial homorganic (sic. — Y. M.) vowels, e.g., Shughni *yast* 'it is, it is present' < **asti*.
2) Consonant metathesis in -CC with C1 weakened and/or C2 strengthened: Shughni *waxt* 'time' < Arabic *waqt*.
3) Final insertion of unstressed -i after postconsonantal sonorants in borrowings (except in Sarikoli): Shughni *fikri* 'thought, idea' < Arabic *fikr*.
4) Epenthetic short i (= /ɪ/) in initial etymological CC-, which are consequently changed to /CiC/ [C(i)C]: Shughni *v(i)rōd* < **brātar-*.

After that, Edelman notes that syllable duration which is different from vowel duration played a role in the history of the vocalic systems, though synchronically is not relevant.

Edelman mentions that Shughni-Rushani words are characterized by 'anlaut articulatory intensity', which means that word-initial CV- is optimal (therefore, initial consonants are articulated with more strength, and prothetic consonants are often inserted). As for word-final syllables, she mentions that the major part of Shughni-Rushani words is consonant-final. A table with average occurrences for 100 words (measured five times) is given below.

Coda	Shughni	Rushani	Khufi	Bartangi	Sarikoli
no coda (-V)	30	23.4	22.8	18.4	20.4
-C	56.4	57.2	57.4	61	60.4
-CC	15.4	18.6	19.6	20.4	17.6
-CCC (verbal 3SG forms and PST stems)	4/500	1/500	1/500	1/500	3/500

Table 1: Syllable structure in the languages of the Shughni-Rushani subgroup. Measured are average occurrences of syllables for 100 words (in five samples). Adopted from [3, p. 213]

Precisely, in Shughni, Rushani, Khufi and Roshorvi, final vowels are usually /aː ɪ ʊ/ (not /a i u/); in Bartangi, /i u/ 'are not rare' in this position.

Nominal stems are usually di- or trisyllabic (simple); compounds and suffixed stems are of three to four syllables (complex). Most of the affixes are monosyllabic (disyllabic are scarce).

In [4], it is stated that Pamir languages have C- or V-initial words. Wakhi and Ishkashimi are said to permit syllables like CCCV; such syllables, according to Payne, are rare in Yazghulami (this is in line with [5, p. 19]) and not present in the Shughni-Rushani subgroup. Payne claims that the latter 'prefers single-consonant onsets except when an unstable vowel is omitted' (p. 427). As for syllable-final position, it is claimed that in all Pamir languages, there can be codas with up to three consonants.

To my knowledge, the only source in which hiatus in Shughni is discussed is the grammar of the Bajuwi dialect by D. Karamshoev. He claims that hiatus is, as a rule, not allowed in the Bajuwi dialect and there are two ways of breaking it: either /j/ is inserted or a vowel merger occurs. The choice of either

means is influenced by the type of vowels and speech register (for example, vowel merger is frequent in fast speech). Then, Karamshoev defines three types of hiatus [6, p. 60]:

1) stem-final + suffix vowels;
2) word-medial vowels;
3) word boundary hiatus.

The first type does not allow hiatus at all, e.g., *garðaː=jaːm χud* 'we ate a flatbread', where /j/ is a hiatus breaker. For the second type, hiatus is possible. Hiatus is observed most often in borrowings pronounced by the intelligentsia and young people, who have mastered Standard Tajik pronunciation (p. 62). In addition to /j/, /w/ and /h/ can be used to eliminate hiatus word-medially: Tajik /mahin/ 'thin' > Bajuwi *majin*, Tajik /muovin/ 'deputy' > Bajuwi *mɔwɪn* or *mʊhɔwɪn*. Lastly, if vowels on the word boundary are pronounced without a pause, they can be merged: *jaː pɛrðøn sɪfaːd* 'She climbed up the hearth', where *pɛrðøn* < *pɪ* 'up' + *arðøn* 'hearth' (p. 63).

2 Studying Shughni phonotactics by means of *pamiri.online*

Among other resources, *pamiri.online* [7] hosts an online dictionary of several Pamir languages, including Shughni. Apart from the original entries that were added by the website's team, there are Shughni lexemes coming from three-volume Karamshoev's dictionary [8–10] and Zarubin's dictionary [11]. Over 25,100 unique forms of Shughni lexemes were obtained from *pamiri.online*; words which were longer than 20 characters or contained a space or Cyrillic symbol were excluded from the sample since, most likely, they either contained a markup mistake or were complex verbs and idioms, which are stored as single forms in the website's database. Morphemes were also excluded from the sample when I studied Shughni words. After that, I changed the Latin symbols used on *pamiri.online* to those of the IPA [12].

2.1 Syllable structure

Overall, the most common syllables which can stand as words are CVCC (1,333 words found) and CVC (1,295). Examples are *tarð* 'fight!' or *xuvd* 'milk' and *bɪʃ* 'udder' or *daːk* 'give!' respectively. CVCC is so common because of the verb stems, often ending in CC; this is also the case for CVCCC. CVCCC (155) and CV (104) are found significantly rarer. Examples of CVCCC are solely verb forms such as *ʁiɣʤd* '(a dog) whines'; as for CV, they are either short verb forms such as *ða* 'give!', or function words such as *ʦa* 'what?', or loanwords such as *ku* 'mountain' and *ʃɔ* 'shah', which were borrowed from Tajik /kuh/ and /ʃoh/ respectively.

Finally, VCC (52) and VC (46) form the last group of relatively common syllable structures. VC is exemplified mainly by function words such as *ʊz* 'I', or onomatopoeic words like *ʊf* 'oh!' and *aʁ* 'scream!', or loanwords like *aːl* 'solution' (Tajik /hal(ː)/). VCC, on the other hand, is primarily characteristic of loanwords like *aːkt* 'deed, document' (Russian /akt/) and *alq* 'throat' (Tajik /halq/), though a few onomatopoeic verb forms also have such structure, e.g., *aʁd* 'scream.3SG.PST'.

All of the remaining syllables are marginal. Data concerning them are summarized in Table 2.

The distribution of Shughni syllables capable of standing alone as words is shown in Figure 1.

Syllable	Occurrences	Examples
CCVC	9	*v(ɪ)rɔd* 'brother', *w(ɪ)zøn* 'know.NPST' and some other words with optional epenthesis
CCVCC	8	*d(ɪ)raχt* 'tree', *d(ɪ)rɤxt* 'rough' and some other words with optional epenthesis
V	7	*e/a/aːɔ* 'hey!' and some other function words
VCCC	3	forms of *angtow* 'cry (about a baby)'
CCV	3	*p(ɪ)rɔ* 'front part', *v(ɪ)rɔ* 'brother', *=ndɪ* 'LOC' (technically not a word)

Table 2: Marginal syllables in Shughni

Compared with Edelman's description [3], my data provides some updates. If optimality is based on frequency, then Shughni has at least four (not three) 'optimal structures': CVCC, CVC, CV and (not included by Edelman) CVCCC, which is typical of some verb forms and even more often present as self-standing words than CV. If not, then, it is required to define precisely what is meant by 'optimal syllables'. In any case, there are more attested syllables than it is described by Edelman. In addition to the mentioned V, VC, VCC, CV, CVC, CVCC and CVCCC, some Shughni words have the structure of CCVC (it is also mentioned by Edelman later in the chapter), CCVCC, VCCC and CCV. While marginal, they should not be overlooked by Shughni grammar accounts.

3 Word structure

3.1 Are there syllables in Shughni?

Identifying syllables is easy when they are words themselves. In other situations, especially when in word-medial position, one has to make certain assumptions regarding principles of syllabification. However, it is sometimes the case that there is no one 'right' way to syllabize the word. This may happen due to the interspeaker variability (e.g., two speakers may have slightly different sonority scale as exemplified in [13, p. 256]), various contexts (e.g., whether a pause follows the word) or even because syllable is just not relevant for the phonology of particular languages. S. V. Kodzasov and I. A. Murav'jova suggested that sometimes there are no phonological criteria of syllabification (such languages are to be called 'wave languages', examples are Russian and Georgian) [14, p. 458–459, 15]. Moreover, there is no generally accepted definition of the syllable [13, 16, p. 3]. As to whether Shughni is a wave language, more research is required. It is clear that some phonetic parameters are dependent on the concept of syllable, e.g., only stressed syllables are heavily aspirated (see [17–19]).

All these points considered, I will discuss only the most tangible aspects of syllabification: word-initial onsets, word-final codas and intervocalic consonants. In the absence of clear evidence of syllable boundaries, the codas of word-initial syllables as well as the onsets of word-final syllables are may well be onsets and codas respectively.

3.2 Word-initial onsets

Most common are word-initial CV (21,691) syllables, significantly less observed is the absence of the onset (2,300); onset CC clusters are marginal (41). Examples: *baːs* 'enough!', *awqɔt* 'provisions', *s(ɪ)tɪrɛʤ* 'female (of species)'.

Attested CC- onsets are /bl/ (*b(ɪ)lisak* 'some insect'), /kl/ (*k(ʊ)lub* 'club'), /fj/ (*f(ɪ)jak* 'small shovel'), /st/ (*S(ɪ)taːlɪnɔbɔd* 'Stalinabad, former name of Dushanbe'), /wz/ (*w(ɪ)zøn* 'know.NPST'). While some of these clusters obey the Sonority Sequencing Principle (/bl kl fj/), it is not true for all (/s/ and /w/ are more sonorous that /t/ and /z/ respectively).

When a word starts with a vowel, it is most likely /a/ (1,398 words found), then go /ɪ/ (372), /ɔ/ (302) and /ʊ/ (170); only 60 words beginning with /aː/ are found (primarily borrowings such as *aːmaq* 'fool' from Tajik /ahmaq/ or *alː* 'solution' from Tajik /hal/). Initial /e/ (38), /ø/ (14), /i/ (9), /u/ (7) and /ɛ/ (4) are marginal. Examples are *etɪborɪ* (or *aːtɪborɪ*) 'trust; hope' (Tajik /e(ʔ)tibor/), *(y)eb* 'sin, guilt' (Tajik /(ʔ)ajb/), i.e., there is a variant with a prothetic /j/ or substituting /aː/; *øm* 'illiterate' (Tajik /(ʔ)om(ː)/), *(w)øq* 'nausea', *ø* 'hey!', *øn* 'yes'; *(j)id* 'celebration' (Tajik /(ʔ)id/), *(j)iɣd* 'gone numb' and some Bajuwi words; *ubɔl(a)ʤɪn* (*ɪbɔl(a)ʤɪn*) 'pathetic', *(w)uvd* 'seven', *ugʊl* 'coal' (Russian /ˈugolʲ/), i.e., either in loanwords or with an alternative; *ɛzʊmɪ* (*ɪzʊmɪ*) 'women's trousers', the three remaining examples were either morphemes (Bajuwi *-ɛrʤ*) or entries for sounds.

While there is no natural class that could generalize these findings, it is interesting that all of the traditionally short vowels (/a ɪ ʊ/) are more frequent word-initially than their long counterparts (/aː i u/). It is, though, not correct that the traditional terms (short vs. long) better describe the observed pattern since /ɔ/ is traditionally long and at the same time is significantly more frequent than short /ʊ/. See Figure 2.

As for word-initial consonants, every phoneme of Shughni is attested in this position. Most common are (in descending order of frequency) /b p s k t m n χ d/, which are found at the beginning of over a

thousand words. On the contrary, /ð ʒ ʤ ɣ θ h/ are found in the onsets of less than 250 words, the latter three phonemes—in less than a hundred.

It is striking that practically all bilabial phonemes are so common in word-initial position. For /b/, it is true partly due to several /b/-initial affixes (*ba-*, *be-*, etc.); the same can be noted about /p/ (*paɾ-*, etc.). Velars except /k/, uvulars except /χ/ and labiodentals and dentals are less common, the latter in particular. Overall, stops are the most preferred onsets while fricatives (except for /s/ and /χ/) and affricates are less typical. See Figure 3.

3.3 Word-final codas

Most common are word-final VC (14,321) syllables, significantly less observed is the absence of the coda (5,783), and -CC codas (4,672). Word-final VCCCs are relatively rare (266) and found in verb forms. Examples: *vɪɾɔd* 'brother', *aɾuzaː* 'daily; everyday', *dayʤ* 'stitch', *anʤaftst* '(it) will start (about snow, rain, etc.)'.

All in all, the observed distribution of word-final syllable coda types resembles that reported by Joy Edelman (see Table 1).

The distribution of word-final vowels is given in Figure 4. /ɪ/ is by far the most common vowel in this position with 3,085 examples in the sample. Next is /aː/ (384), followed by significantly less frequent /ɔ/ (494) and /a/ (355). The remaining /i/ (66), /ʊ/ (51), /ɛ/ (47), /u/ (43), /e/ (38) and /ø/ (24) are marginal. Examples are *etɪborɪ* 'trust; hope', *baʧ(ː)aː* 'child', *lɪxɔ* 'jaw', *dɪga* 'other'; *maki* 'uncle', *ʤɔdʊ* 'sorcerer', *sarfaɾmøndɛ* 'commander-in-chief', *ruparu* 'opposite', *xʊmne* 'tomorrow', *galø* 'silly'.

According to [3], final vowels are usually /aː ɪ ʊ/ and not /a i u/. As shown, this contradicts the data from *pamiri.online*. /ʊ/ is as marginal as /u/ while /a/, which was excluded from the vowels typical of final position, is rather frequent word-finally. It would also be fair to include /ɔ/ in the list of final vowels. As for /ɛ/, it is in fact the most common final vowel. The reason is that /ɪ/, at least in the speech of young Shughni speakers from Khorugh, is realized as [ɛ] word-finally. This may lead to the conclusion that word-final /ɪ/ has been replaced by /ɛ/, see also [19, p. 189].

Word-final VCs most often have alveolars /t/ (2,374), /n/ (2,043), /ɾ/ (2,012). Quite close to this group are /k/ (1,801), /d/ (1,720) and /ʤ/ (1,611). On the contrary, all fricatives, namely, /x/ (378), /θ/ (332), /z/ (311), /s/ (294), /χ/ (215), /v/ (207), /f/ (180), /ʃ/ (158), /ð/ (152), /ʁ/ (134), /ʒ/ (110), /ɣ/ (80) and /h/ (20), are relatively rare in word-final position. While bilabial /p/ and /b/ were most common word-initially, they are not typical of word-final position. See Figure 5.

As for -CC coda clusters, 211 consonant combinations are attested. Most common are /st/ (468) and /nd/ (438). It is typical for a -CC coda to end in either plosive (2,750) or affricate (1,547); significantly less common are terminal fricatives (315) and nasals (159). Terminal approximants and taps in word-final -CC codas are marginal. The only terminal fricative in Figure 6 is /ð/ in /ɾð/, observed in 45 words such as *angurð* 'vine' and *gaːɾð* 'turn!'. As for the only terminal nasal in the same figure, it is /m/, which is part of word-final /ɾm/ (37) in words like *ʧarm(ɪ)* 'skin, hide' (Tajik /ʧarm/) and *gaːɾm* 'warm' (Tajik /garm/), which are borrowings from Tajik; however, there are Shughni lexemes not loaned from Tajik which end in such cluster, e.g., *vɪdɪɾm* (*ðɪdɪɾm*) 'besom'.

As for the sonority sequence, it is true that in most cases (4,527 vs. 271), the first consonant (C1) is more sonorous than the second (C2) provided that the sonority scale is like (2):

(2) *Shughni sonority scale, cf. [16, p. 7]*

plosives < affricates < fricatives < nasals < tap < approximants < vowels

When it comes to word-final -CCC codas, they generally follow the patterns of the -CC codas. Overall, there are 66 consonant combinations attested, /ftst/ (16), /χtst mbt ɾθt/ (13), /ɾðd fst/ (11) being most common. See Figure 7.

In terminal position, there is either a plosive (164) or affricate (98); terminal nasals (5) are rare. Notably, neither fricatives nor taps and approximants are attested at the end of word-final -CCC codas.

From the perspective of sonority sequencing, most common are word-final CCC clusters which are organized following the ascending sonority pattern (C1 > C2 > C3, C3 being least sonorous). Nevertheless, in 75 words, which is almost half of the C1 > C2 > C3 pattern's occurrences, C2 is less sonorous than C3 (or is equal in this regard). Further, the inversed pattern, when C1 is less sonorous

than C2 or is equal but C2 is more sonorous than C3, is observed in 28 words. Finally, only 4 sequences do not obey the sonority sequencing principle at all.

3.4 Intervocalic clusters

Overall, hiatus is dispreferred in Shughni (268 cases attested). It is most typical of borrowings (e.g., *mʊaɾiɾ* 'editor' < Tajik /muharːir/) and abstract lexemes (*beɪʃtɪjɔ* 'not feeling good'), especially with prefixes like *be-* 'without'. Normally, up to three intervocalic consonants are found, few cases of intervocalic CCCC being an exception (*vɛɣʤnzɔɾ* 'birch grove', *ʃitɔrrθkzɔɾ* 'rhubarb thickets'). See Figure 8.

Most common is intervocalic /ɾ/ (2,658), followed by sonorants /l/ (1,491), /m/ (1,008) and /n/ (986)/. On the contrary, /ɣ/ (115), /ʒ/ (109), /h/ (84), /ʤ/ (53) and /θ/ (48) are not typical of this position. See Figure 9.

Shughni consonants of each manner of articulation are met in intervocalic position. Most common are plosives (4,659), though all of the other categories except affricates are found in more than 2,000 words. See Figure 10.

As for intervocalic CCs, 640 consonant combinations are attested. Most common are /nd/ (517), /st/ (299), /ng/ (270), /mb/ (238), /ɾm/ (229), /ɾð/ (222). Those clusters that are found in more than 100 words are given in Figure 11.

In respect of manner of articulation, the most common C1s are nasals and fricatives; the most common C2s are plosives. See Figure 12.

As for Sonority Sequencing Principle, the major part of the intervocalic CCs obeys it (7,874) while in 2,352 cases, C1 < C2, and in 1,846 words, C1 = C2 in respect of sonority.

Intervocalic CCCs are of 405 different types, though only seven of them are observed in at least 10 words: /ngt/ (19), /ɣʤd mbt/ (12), /ɾθt/ (11), /stb stk ɾʤb/ (10). In 570 words, C1 is more sonorous than C2 while C2 is less sonorous than C2 or is equal in this regard. In 244 words, however, the Sonority Sequencing Principle is not defied. See Figure 13 for other patterns.

Finally, CCCCs in intervocalic position are found in few words with the suffix *-zɔɾ* 'place': *vɛɣʤnzɔɾ* 'birch grove' (cf. *vɛɣʤ(n)* 'birch'), *ʃitɔɾθkzɔɾ* 'rhubarb thickets' (cf. *ʃitɔɾθk* 'rhubarb'). Note that *vɛɣʤnzɔɾ* is likely to be pronounced with a syllabic nasal.

The most common geminate consonants in intervocalic position are liquid /l/ (155) and /ɾ/ (83). See Figure 14 for all attested intervocalic geminates.

4 Conclusion

Studying phonotactics by means of computational methods allows for comprehensive and precise descriptions, which are hard to achieve using traditional means of analysis. Describing Shughni phonotactics used to be a difficult task, which is supported by the scarcity of its accounts in the literature. Now that several Shughni dictionaries are available as datasets, it is convenient to apply them for the study of grammar, and sound patterns in particular.

This paper provides an example of such phonotactic research based on data from *pamiri.online*. My findings have (dis)proved some of the previous claims regarding Shughni phonology and also brought about new facts which, as I hope, will be useful in creating a new grammar description of Shughni.

Acknowledgements

I am grateful to the two anonymous reviewers for their valuable comments. The publication was prepared within the framework of the Academic Fund Program at HSE University (grant No 22-00-034 'Field and Corpus Research on the Languages of the Shughni-Rushani Group').

References

[1] Olson K. Shughni Phonology Statement. SIL International, 2017.
[2] Edelman D. (Joy) I., Dodykhudoeva L.R. Shughni // The Iranian languages. 1st ed. / ed. Windfuhr G. London and New York: Routledge, 2009. P. 787–824.

[3] Edelman D. (Joy) I. Shugnano-rushanskaja gruppa [Shughni-Rushani group] // Jazyki mira: Iranskiye yazyki. III. Vostochnoiranskiye yazyki. Moscow: Indrik, 1999. Vol. III. P. 225–242.
[4] Payne J.R. Pamir languages // Compendium Linguarum Iranicarum / ed. Schmitt R. Wiesbaden: Reichert, 1989. P. 417–444.
[5] Edelman D. (Joy) I. Jazgulamskij jazyk [Yazghulami language] / ed. Abaev V.I. Moscow: Nauka, 1966.
[6] Karamshoev D. Badzhuvskij dialekt shugnanskogo yazyka [Bajuwi dialect of Shughni]. Dushanbe: Izdatel'stvo AN Tadzhikskoj SSR, 1963.
[7] Makarov Y., Melenchenko M., Novokshanov D. Digital Resources for the Shughni Language // Proceedings of The Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia within the 13th Language Resources and Evaluation Conference. Marseille, France: European Language Resources Association, 2022. P. 61–64.
[8] Karamshoev D. Shugnansko-russkij slovar' [Shughni-Russian dictionary]. Moscow: Izdatel'stvo "Nauka," 1988. Vol. 1.
[9] Karamshoev D. Shugnansko-russkij slovar' v tr'ox tomax [Shughni-Russian dictionary in three volumes]. Moscow: Nauka, 1991. Vol. 2. 615 p.
[10] Karamshoev D. Shugnansko-russkij slovar' [Shughni-Russian dictionary]. Moscow: Izdatel'skaja firma "Vostochnaja literatura" RAN, 1999. Vol. 3. 568 p.
[11] Zarubin I.I. Shugnanskije teksty i slovar' [Shughni texts and dictionary]. Moscow, Leningrad: Izdatel'stvo AN SSSR, 1960.
[12] Nolan F. et al. Handbook of the International Phonetic Association. New York: Cambridge University Press, 1999.
[13] Ladefoged P., Johnson K. A Course in Phonetics. 7th ed. Stamford, USA: Cengage Learning, 2014. 352 p.
[14] Kodzasov S.V., Krivnova O.F. Obschaja fonetika [General phonetics]. Moscow: Russian State University for the Humanities, 2001. 592 p.
[15] Kodzasov S.V., Murav'jova I.A. Slog i ritmika v al'uterskom jazyke [Syllable and rhythm in Alutor] // Publikatsyi OSIPL MGU. Moscow, 1980. Vol. 9.
[16] Easterday S. Highly complex syllable structure: A typological and diachronic study. Berlin: Language Science Press, 2019.
[17] Makarov Y. Plosives in Shughni. Moscow: HSE University, 2022.
[18] Makarov Y. Realizatsyja nachal'nykh vzryvnykh v shugnanskom jazyke [Realisation of the initial plosives in Shughni]. Moscow, 2022.
[19] Makarov Y.Y., Plungian V.A. K opisaniju shugnanskoj morfonologii: predvaritel'nyje zamechanija [Towards the Description of Shughni Morphonology] // «…Vperjod i vverx po lestnice zvuchashhej». Sbornik statej k 80-letiju Ol'gi F'odorovny Krivnovoj. Moscow: Buki Vedi, 2023. P. 178–193.

Appendix

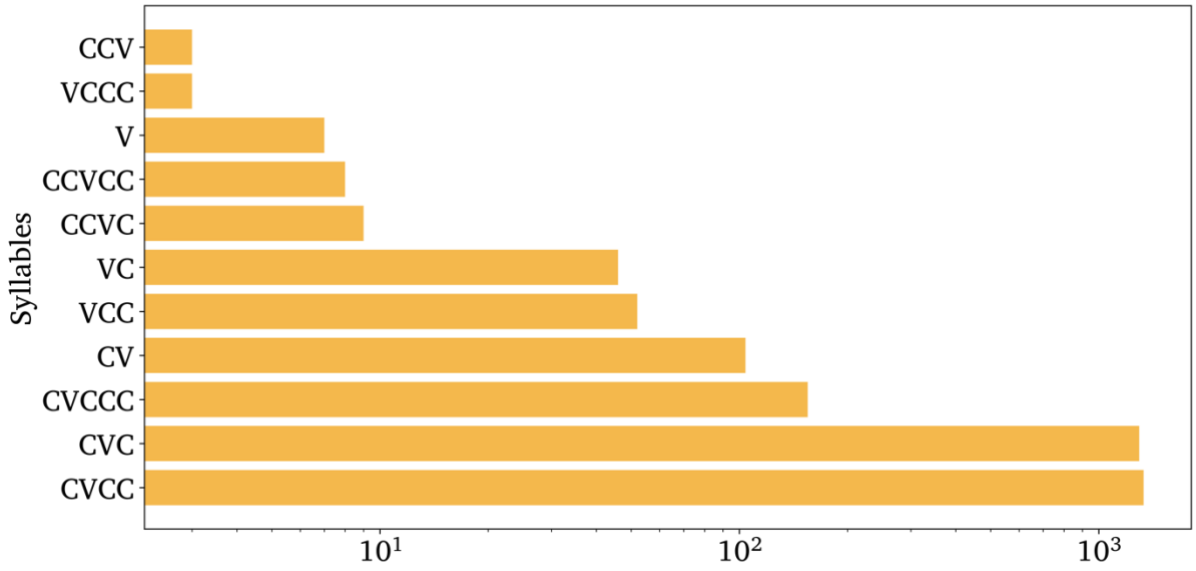

Figure 1: Distribution of self-standing syllables in Shughni. The horizontal axis is a log scale (in favour of visibility) and does not reflect real number of occurrences.

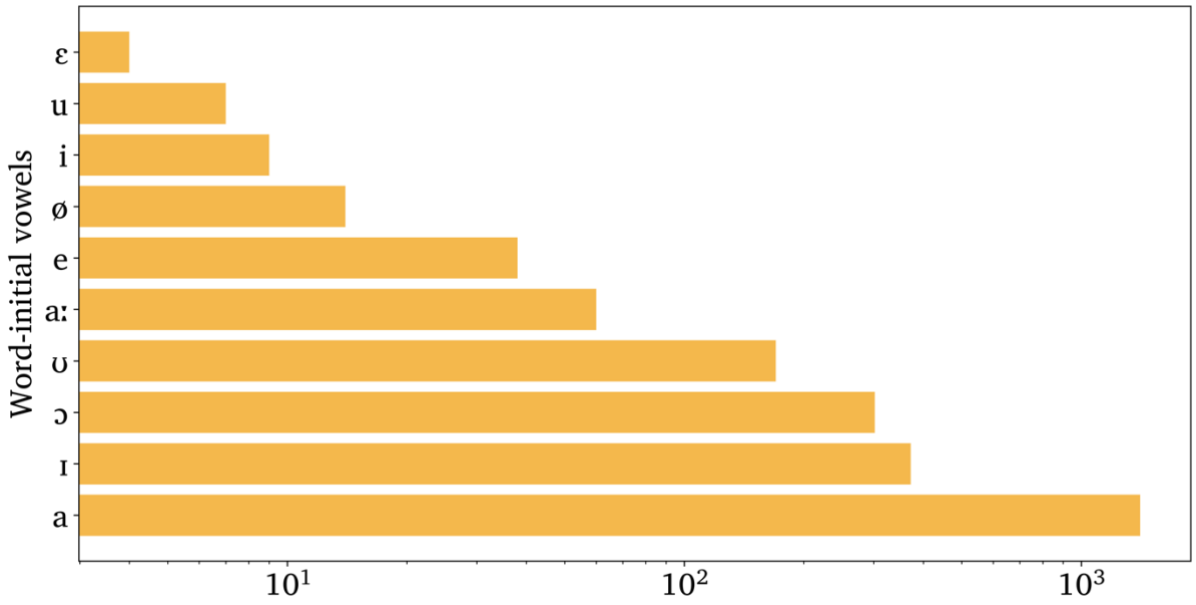

Figure 2: Distribution of word-initial vowels in Shughni. The horizontal axis is a log scale

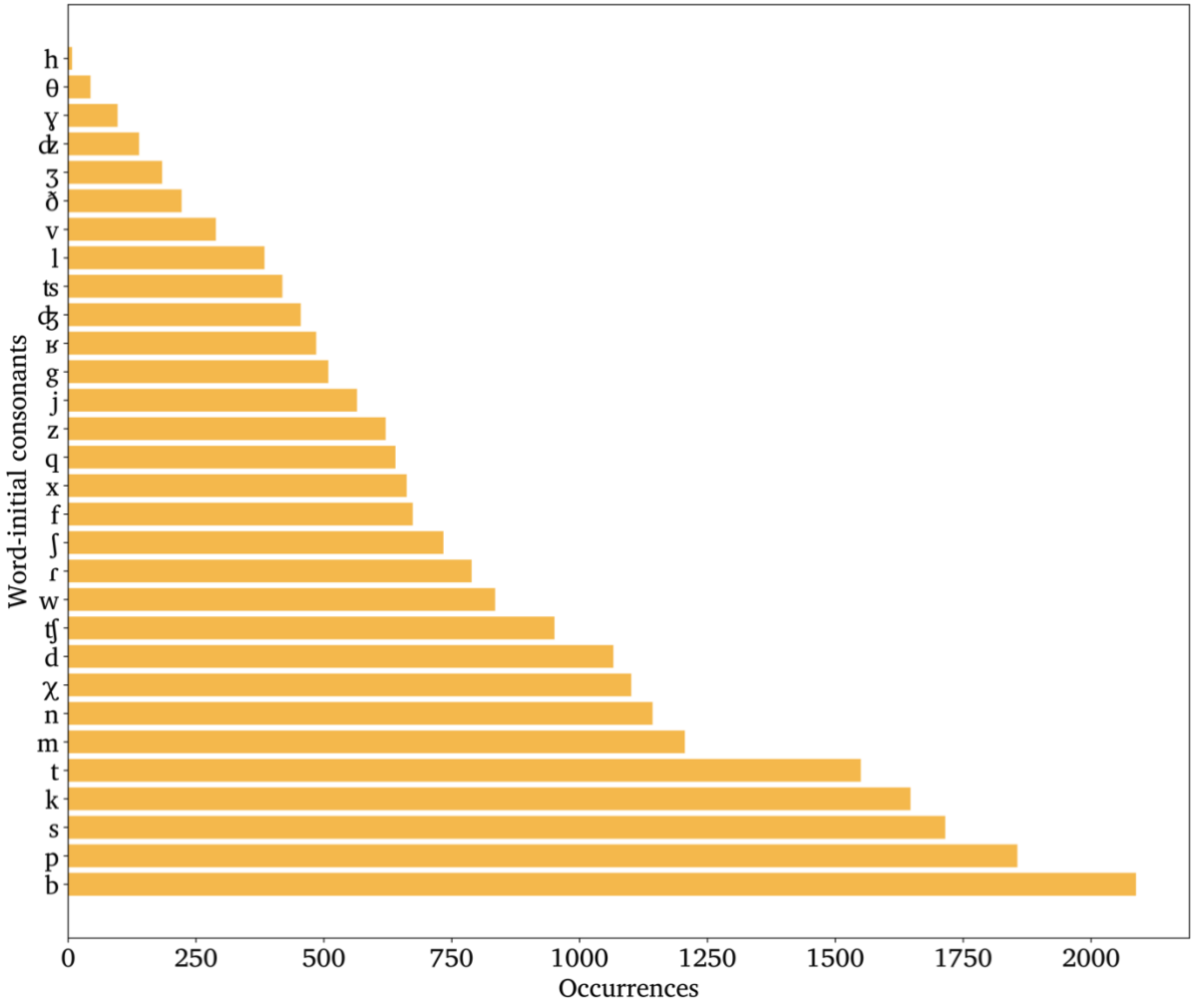

Figure 3: Distribution of word-initial consonants in Shughni

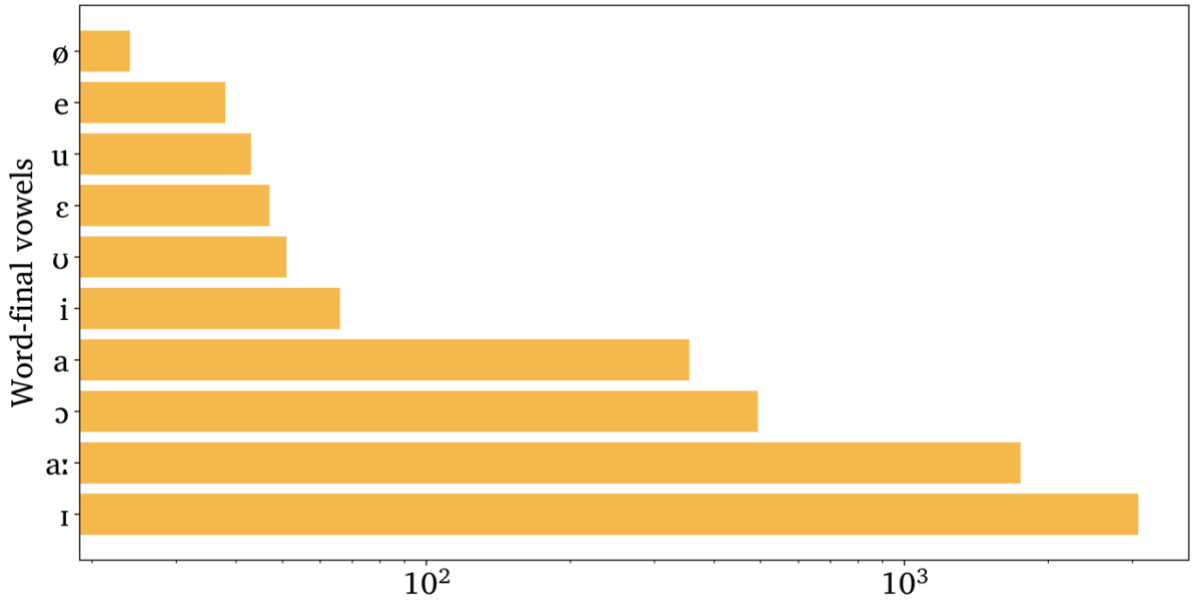

Figure 4: Distribution of word-final vowels in Shughni. The horizontal axis is a log scale

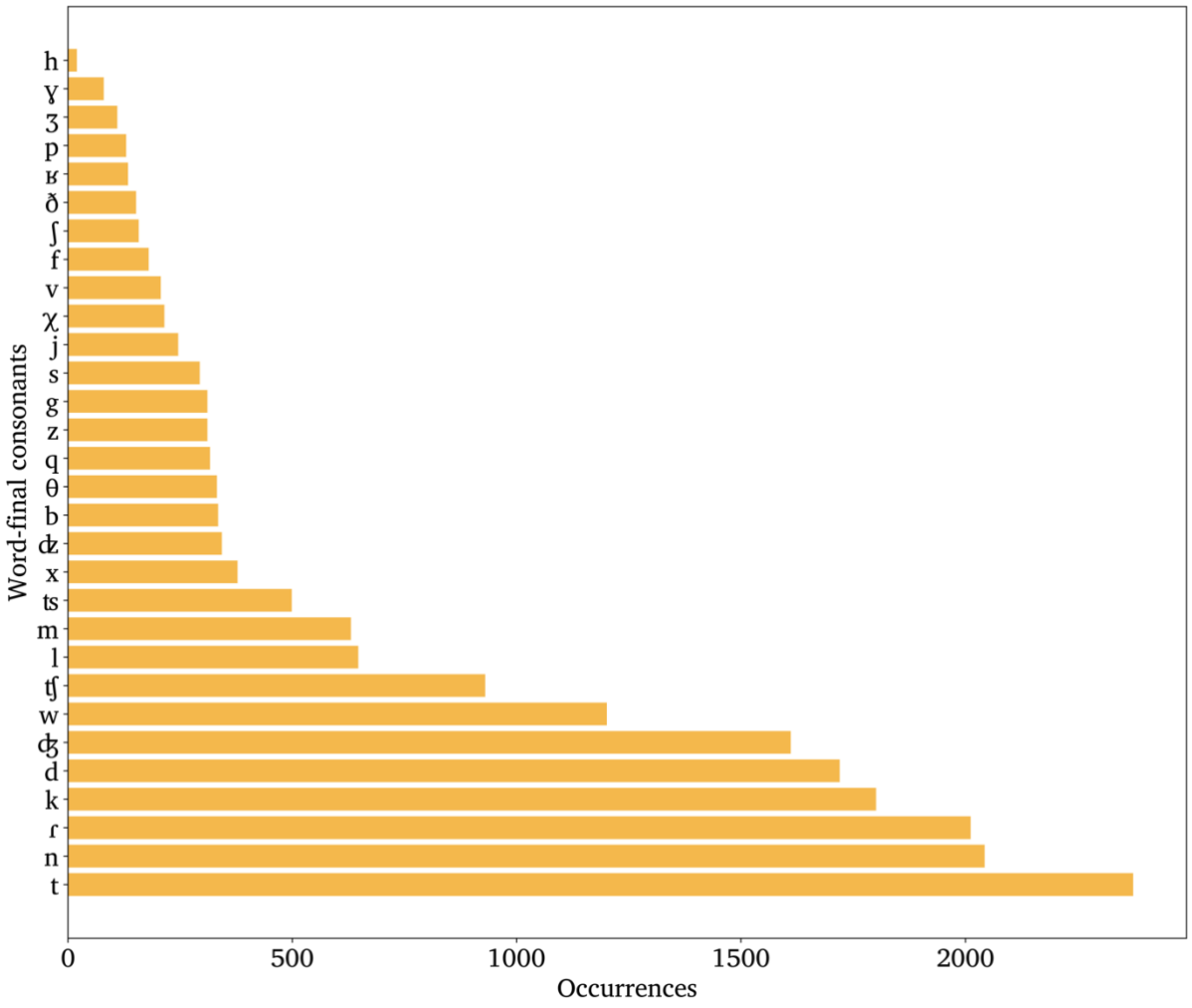

Figure 5: Distribution of word-final consonants in Shughni

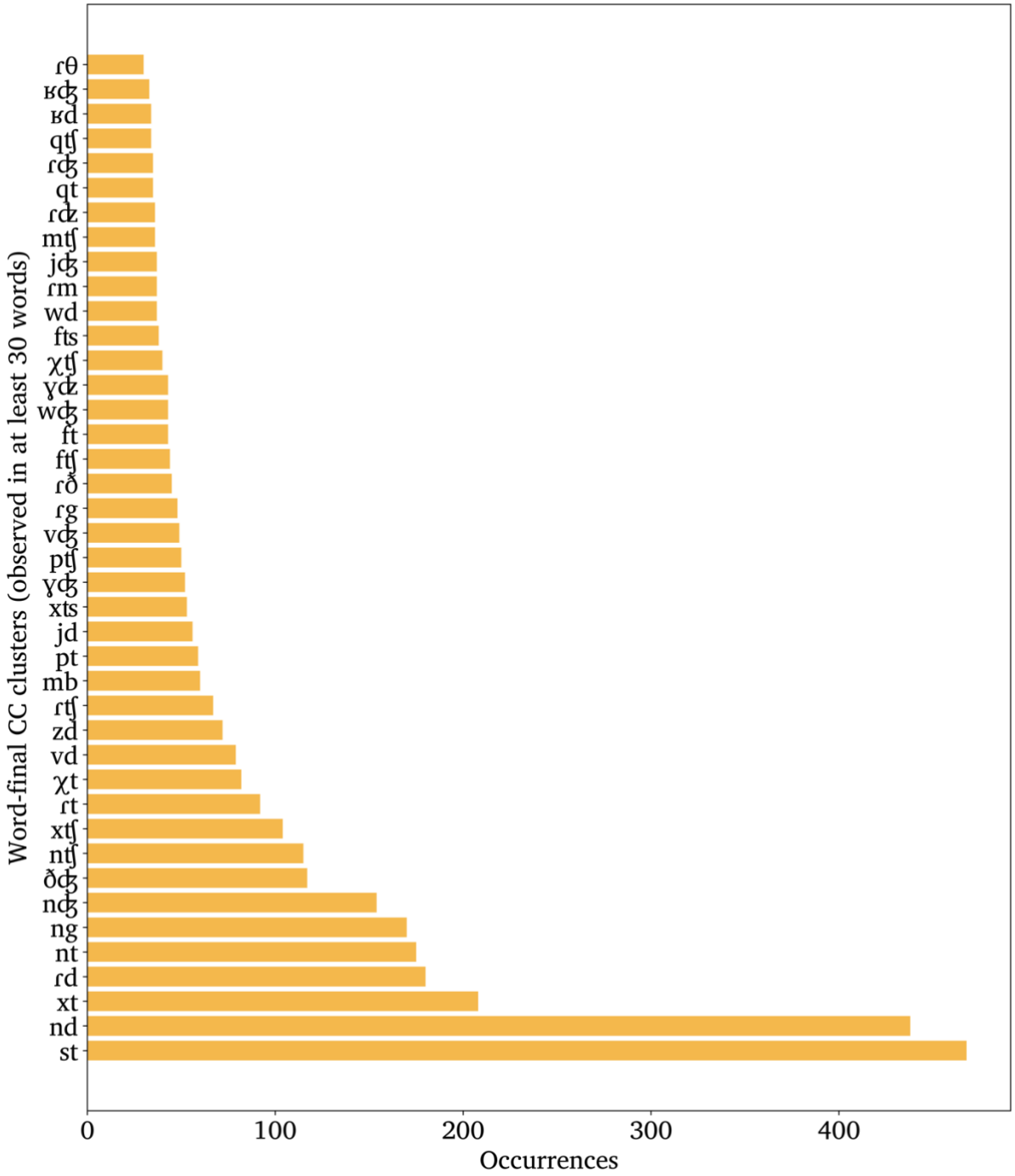

Figure 6: Distribution of word-final CC clusters in Shughni. Only the clusters observed in at least 30 words are included

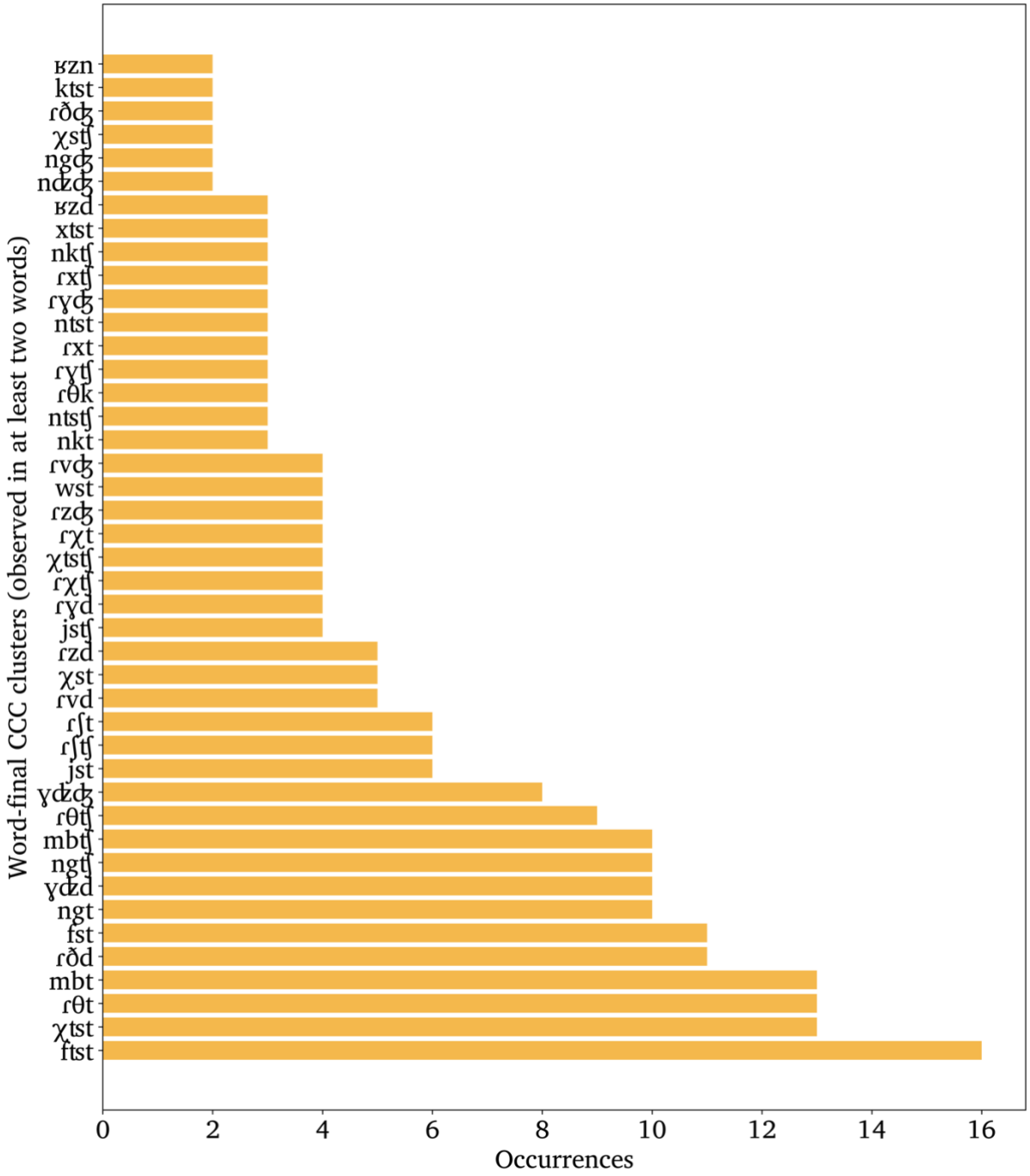

Figure 7: Distribution of word-final CCC clusters in Shughni. Only the clusters observed in at least two words are included

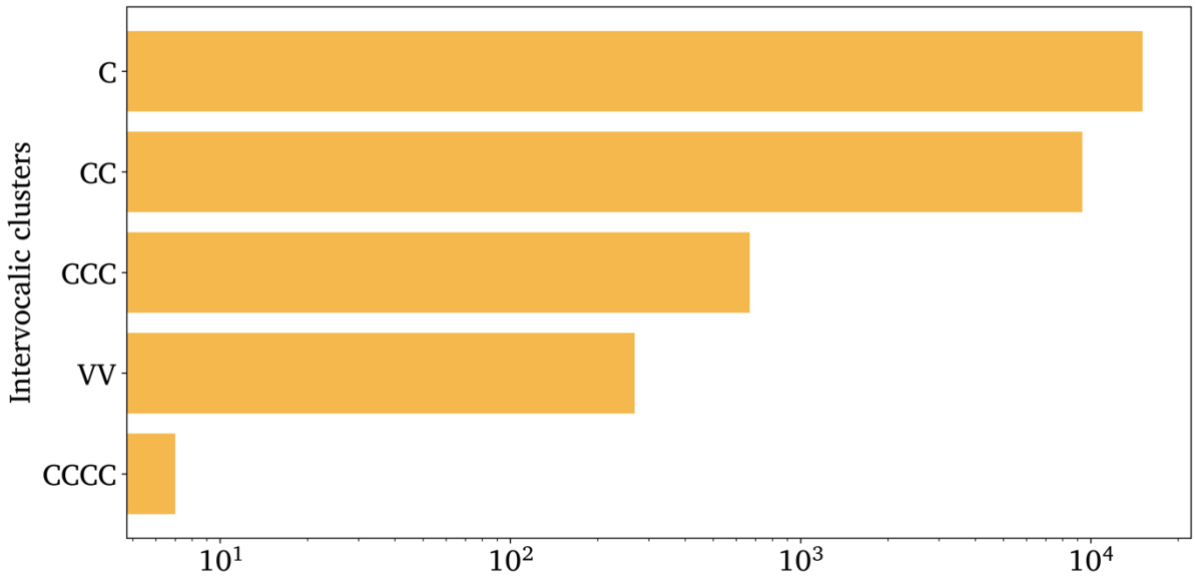

Figure 8. Distribution of intervocalic consonants in Shughni. The horizontal axis is a log scale, *VV* stands for hiatus

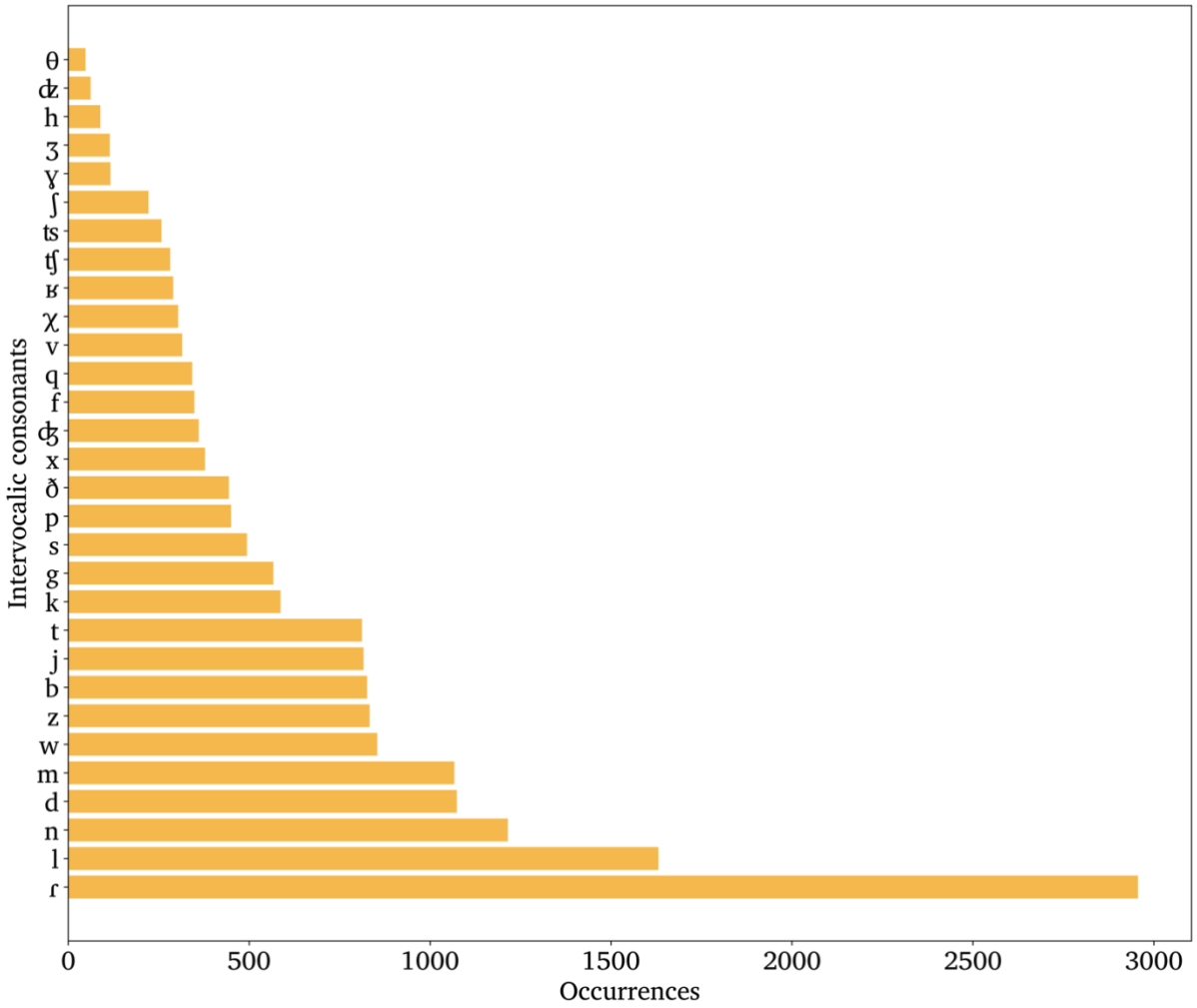

Figure 9: Distribution of Shughni intervocalic consonants

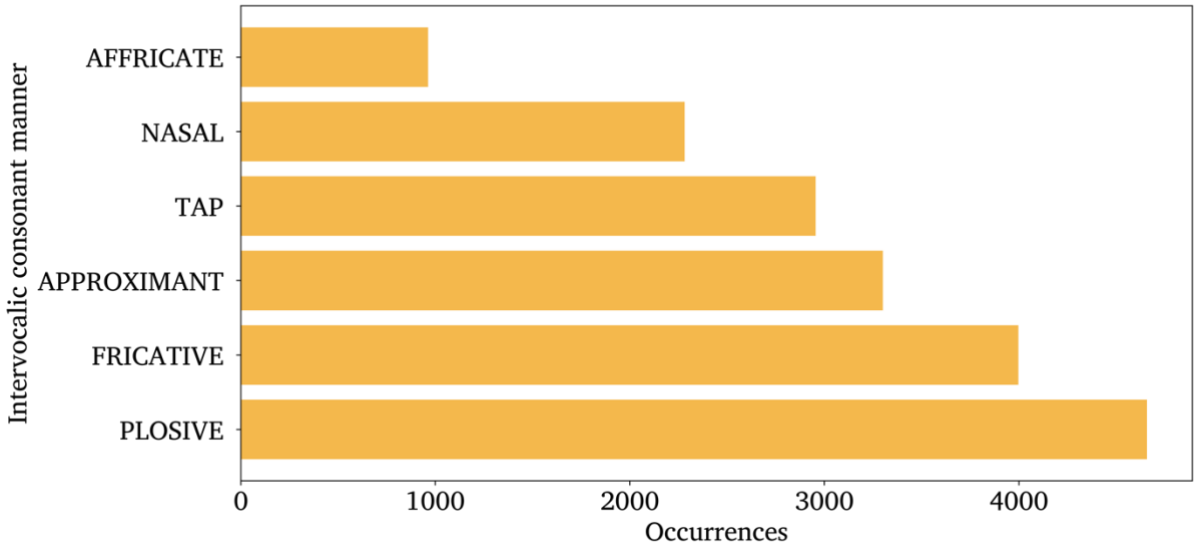

Figure 10: Manners of articulation of intervocalic consonants

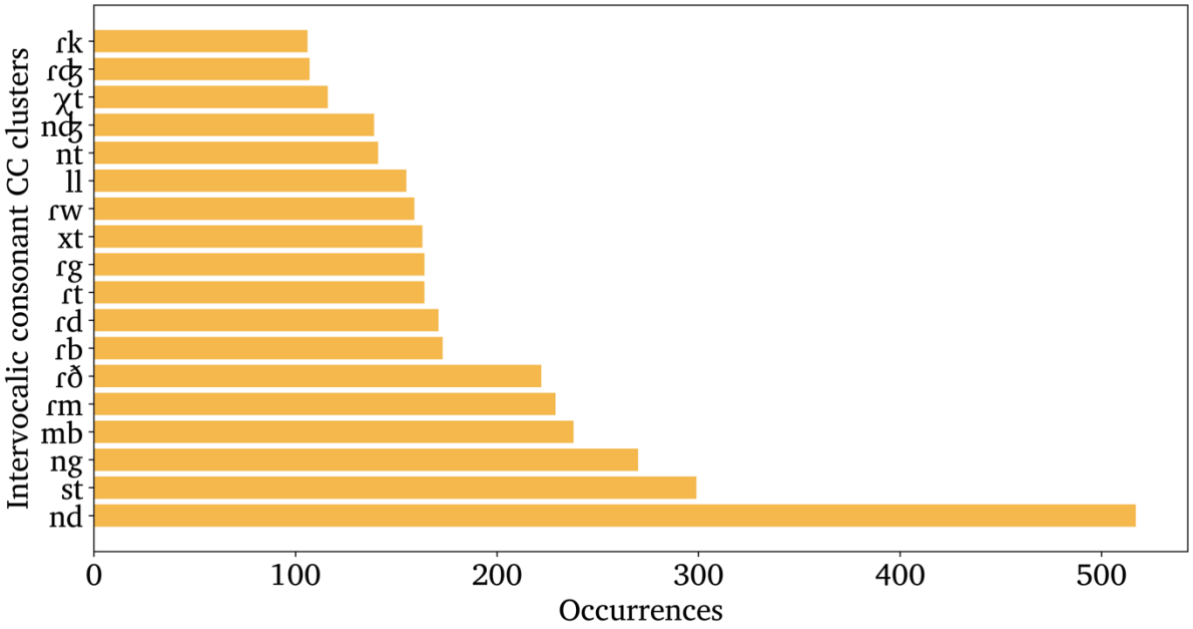

Figure 11: Intervocalic CC clusters in Shughni (found in more than 100 words)

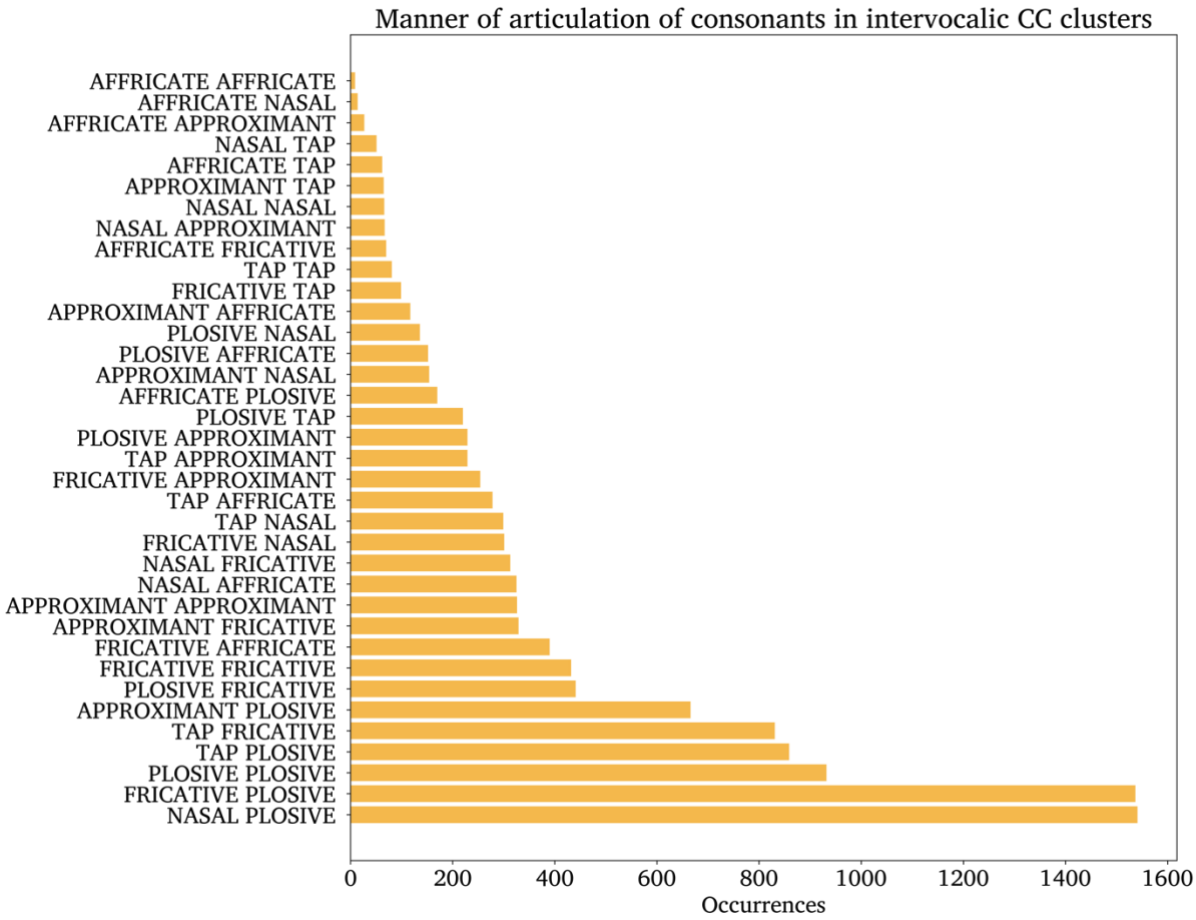

Figure 12: Manner of articulation of consonants in intervocalic CC clusters in Shughni

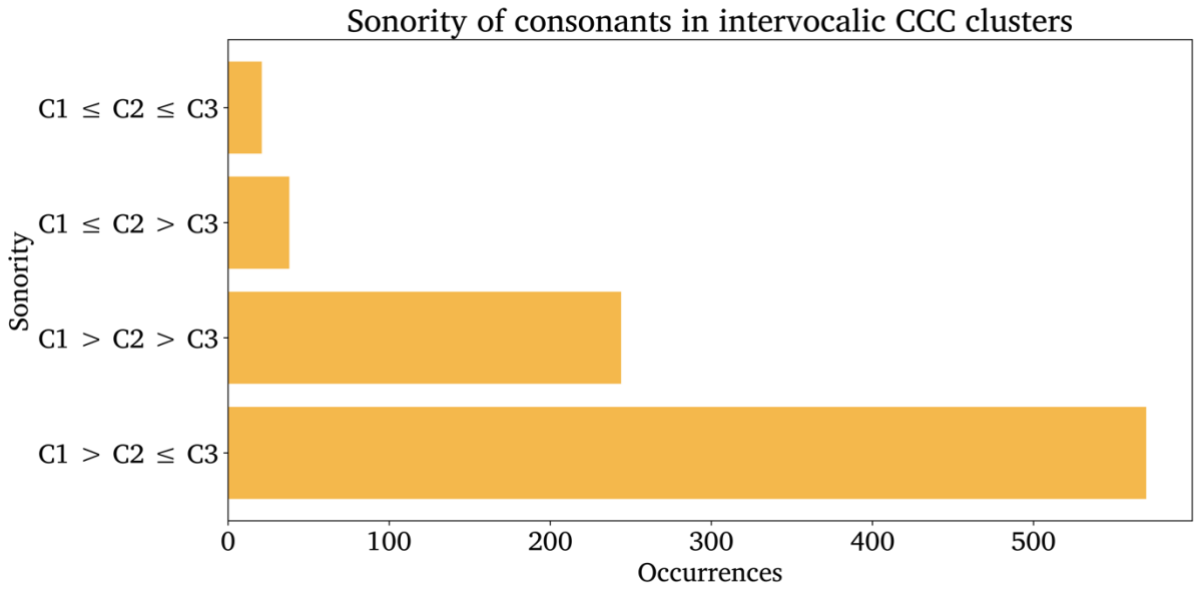

Figure 13: Sonority sequencing in intervocalic CCC clusters in Shughni

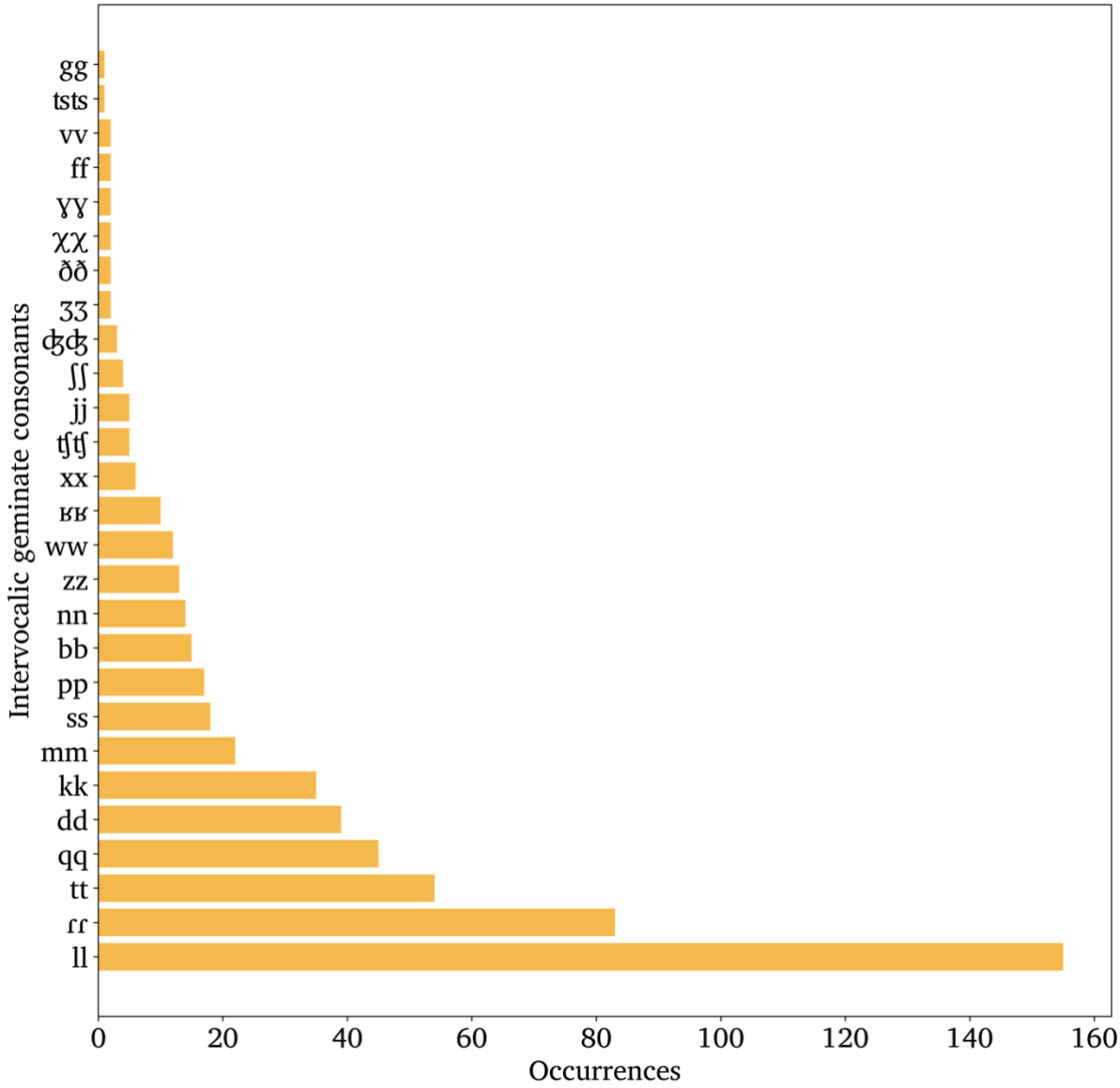

Figure 14: Intervocalic geminates in Shughni