

Typology and Phonetic Realization of Final Consonant Clusters in the Tajik Language

PANJIYEV NORMAXMAT
Termez State University

ABSTRACT

This article examines the structural and phonetic features of word-final consonant clusters in the modern Tajik literary language. The study identifies and classifies final clusters according to the number of components and the interaction of obstruents and sonorants. Based on dictionary and grammatical sources, 222 lexical attestations were classified into 125 structural types, including 114 binary, 10 ternary, and one quaternary type. The analysis focuses on positional devoicing, assimilation, fricative realization, and the phonetic behavior of glottal stops and final geminates. The findings show that binary clusters constitute the structural core of the system, while multi-component clusters mainly occur in borrowed and international vocabulary. Although final clusters generally follow the principle of decreasing sonority, lexical and historical factors allow several exceptions. The proposed classification may contribute to comparative phonotactic research, pronunciation lexicography, and speech technology.

Keywords: Tajik language, phonotactics, final consonant clusters, word-final position, consonant combinations, phonological structure, phonetic realization, sonority, sonority sequencing principle, assimilation, final devoicing, articulatory processes, binary clusters, ternary clusters, borrowed vocabulary, phonological patterns

INTRODUCTION

Final consonant combinations constitute one of the domains of the phonological structure of the word in which phonological constraints, articulatory patterns, and the history of the lexical inventory interact particularly clearly. Word-final position intensifies positional processes such as the devoicing of obstruents, assimilation according to place and manner of articulation, and the reduction of individual segments. Therefore, the description of word-final structures requires a distinction between orthographic form, phonemic composition, and normative phonetic realization.

In Tajik linguistics, the consonant system and syllable structure have been described in detail in both classical and modern grammars [4: 15-56; 12: 17-38], while orthographic dictionaries and the current spelling rules provide extensive empirical material without, however, demonstrating which patterns constitute the systemic core [5: 1-400; 3: 1-24].

The theoretical significance of the topic is associated with the problem of the sonority profile of the syllable: sonority decreases from the nucleus toward the periphery, and the coda is characterized by a descending contour [6: 283-333; 8: 243-279]. However, morpheme boundaries, segment licensing, and borrowings create numerous exceptions [15: 1-303; 13: 107-136]. The Tajik material is particularly illustrative because the inherited Iranian lexical stratum coexists with long-established Arabic vocabulary and international terminology. Consequently, the final subsystem combines frequent binary structures (-нд, -рд, -ст) with peripheral multi-component structures (-кст, -нстр).

The aim of the article is to identify the composition and principles of the structural and phonetic organization of final consonant clusters in the modern Tajik literary language: to determine the scope of the inventory, classify final clusters according to the number of components, describe binary models according to the relationship between obstruents and sonorants, and characterize positional phonetic processes. The scientific novelty lies in integrating orthographic, articulatory, and stratificational parameters into a unified system comprising 125

types. The working hypothesis is that final phonotactics is organized at two levels: a stable core of binary clusters and an expandable periphery of multi-component and borrowed forms, with increasing orthographic complexity not necessarily corresponding to increasing phonetic complexity.

MATERIALS AND METHODS

The research material consisted of a revised table of final consonant clusters in Tajik, compiled on the basis of the 2025 academic orthographic dictionary, its 2026 electronic edition, the current Tajik orthographic rules, and the machine-readable Tajik HunsPELL dictionary (68665 entries) [5: 1-400; 3: 1-20; 14]. Grammatical descriptions of Tajik [12: 17-38; 4: 15-56] were used for the interpretation of pronunciation, while the general analytical model was based on studies of phonotactics and syllable structure [1: 1-62; 7: 103-140; 9: 91-129]. The unit of analysis is a type of final sequence rather than an individual word: for example, банд, қанд, панд and баланд illustrate one type, -нд. A total of 222 dictionary attestations were recorded and distributed among 125 types; this indicator reflects lexical attestation rather than corpus frequency.

A final cluster is defined as the maximum continuous sequence of consonants following the last vowel up to the end of the word (дарахт → -хт, пункт → -нкт, министр → -стр, монстр → -нстр); this approach is consistent with the distributional understanding of phonotactics [2: 70-110; 10: 44-78]. Each type was coded according to the number of components, orthographic and phonetic form, participation of voice and noise, manner of articulation, and lexical and stylistic status. In the notation, III denotes an obstruent, C a sonorant, Cм a stop, III a fricative, H a nasal, Л a lateral, Д p a trill, Ап an approximant, and Аф an affricate; ʔ is regarded as a graphic indicator of the historical glottal stop, conventionally transcribed as [ʔ].

The study employed the methods of exhaustive sampling, positional analysis, articulatory analysis (the sequence C₁ → C₂ → ...), distributional analysis, quantitative analysis, and

stratificational analysis, as well as sonority-based interpretation as an explanatory rather than prescriptive principle [6: 283-333; 15: 1-303]. The transcription represents normalized pre-pausal realization: банд [bant], дард [dart], арз [ars], қалб [qalp]; [ʔ] in the forms табъ and манъ marks the normalized glottal stop. A limitation of the study is the absence of an acoustic experiment: the table describes structural types but does not provide data on stop duration, the percentage of devoicing, or regional variation in spoken language.

RESULTS

The quantitative distribution demonstrates a pronounced asymmetry according to cluster length: 114 binary types constitute 91.2 % of the inventory, ten ternary types account for 8.0 %, and one quaternary type accounts for 0.8 % (Table 1). Binarity represents the principal structural norm of Tajik word-final position; multi-component structures are systematically restricted and lexically marked.

Table 1. *Distribution of final clusters according to the number of components*

Number of components	Number of type	Share %	Structural status
2	114	91,2	core
3	10	8,0	periphery
4	1	0,8	extreme periphery
TOTAL	125	100.0	—

Within the binary class, seven macro-models can be distinguished according to the relationship between obstruency and sonority (Table 2). The largest categories are III+III and C+III, with 40 types each (35.1 %); III+C comprises 18 types (15.8 %), while C+C includes 8 types (7.0 %). Combinations involving affricates are rare (C+Aϕ – 4 types, Aϕ+III and Aϕ+C – 2 types each). The equal representation of III+III and C+III reflects two different coda strategies: the accumulation of obstruents and the transition from a more sonorous segment to a less sonorous one.

Table 2. *Macro-models of binary final consonant clusters*

Macro-model	Number	Share, %	Examples
Ш+Ш	40	35.1	-ст, -фт, -хт, -кт
С+Ш	40	35.1	-нд, -рд, -рф, -нк
Ш+С	18	15.8	-бр, -зм, -см, -хр
С+С	8	7.0	-йл, -лм, -рм, -мн
С+Аф	4	3.5	-нч, -нц, -рч, -рц
Аф+Ш	2	1.8	-чз, -чъ
Аф+С	2	1.8	-чм, -чр
TOTAL	114	100.0	—

According to manner of articulation, the most numerous model is Ш(Ш)+Ш(См), comprising 20 types (-ст, -фт, -хт, -кт): the fricative phase of C□ smoothly transitions into the complete obstruction of C□, with both components usually voiceless. The С(Н)+Ш(См) model comprises 10 types (-нг, -нд, -нк, -мб); in -нг and -нк, /n/ assimilates to the following velar, so that панг and банк coincide word-finally as [raŋk]/[baŋk]. The rhotic sonorant forms 14 types with an obstruent (-пв, -пз, -пф, -пб, -пд), encompassing both inherited vocabulary (гарм, сард, мард) and Arabic borrowings (харф, фарз, шарк); the reverse order (-бр, -др, -тр, -хр, -шр) demonstrates a local rise in sonority but is firmly established lexically (сабр, кадр, шахр).

The second component of binary final clusters is less diverse: a voiced or voiceless stop occurs in 50 types, a fricative in 32, rhotic and nasal sonorants in 11 each, a lateral in six, and an affricate in four. Thus, 82 of the 114 types (71.9 %) terminate in a stop or fricative obstruent, confirming the tendency of Tajik word-final position toward a voiceless obstruent outcome. A special subsystem is formed by combinations with ь (-бъ, -мъ, -ръ, -тъ), normalized as [pʔ], [mʔ], [rʔ], [tʔ], and by the geminates -дд, -зз (хадд, қадд, радд, қазз), realized before a pause as long voiceless [t:] and [s:].

Positional devoicing affects final /б, д, з, в, г, ғ, ч/: банд [bant], дард [dart], арз [ars]. A separate regularity is represented by combinations with /к/ before a consonant (-қд, -қл, -қс, -қт, -қф, -қш): the stop is realized as the fricative [χ] – нақд [naχt], вақт [vaχt], нақш [naχʃ]; вақт and дарахт coincide in their final sequence as [χt], although they are orthographically distinct.

Ternary final clusters are represented by ten types (-кст, -нкт, -нтр, -ндр, -ктр, -лтр, -стр, -мск, -нск, -рск), predominantly belonging to international terminology (текст, пункт, министр, оркестр, спектр) and toponyms (Омск, Минск, Курск); the only quaternary type, -нстр (the C+III+III+C structure), is represented by the word монстр; none of these types forms a broad inherited lexical series.

The functional core of the system consists of types attested by several commonly used words: -нг (ранг, чанг), -нд (банд, баланд), -рд (дард, сард), -рз (арз, фарз), -ст (даст, рост), -хт (бахт, дарахт), -хр (шахр, мехр). Type productivity does not always coincide with phonetic simplicity: -хр and -шм are stable despite a local rise in sonority, whereas formally simple -пс and -пт are restricted to isolated borrowings (гипс, рецепт, концепт).

DISCUSSION

The findings confirm the typological expectation of decreasing sonority toward the end of the word: forty C+III types and a substantial proportion of III+III types exhibit a descending profile. However, 18 III+C types and a number of multi-component combinations demonstrate that the sonority sequencing principle is not absolute: universal preference interacts with language-specific licensing of particular structures [6: 283-333; 15: 1-303].

For Tajik word-final position, it is productive to distinguish three zones: the central zone (inherited and long-established vocabulary, including the Arabic-Persian -бр, -др, -зм, -хр), the intermediate zone (bookish and terminological binary types), and the outer zone (multi-component international and onomastic structures). The stability of sequences is ensured by frequent words, morphological transparency, and the prestige of the written form [11: 207-263; 9: 91-129]; a borrowing that has become firmly integrated into the word-formation relations of the language thereby becomes part of the core regardless of its origin.

The status of -нг is particularly illustrative: assimilation of /n/ and devoicing of /g/ lead to the same result, [ŋk], as the

borrowed -HK, providing an example of the neutralization of an entire cluster opposition rather than of an individual phoneme. Similarly, spirantization of /κ/ brings ваKT and дапахT closer together through the final [χt]. Such correspondences are important for lexicography (notations: паHT [paŋk], ваKT [vaχt]), language teaching (from central binary models to borrowings), and speech technologies, where contextual grapheme–phoneme rules are more effective than a list of exceptions.

The material involving Ъ and final geminates requires experimental phonetic verification; further research should involve corpus-based analysis of the frequency of types and tokens, as well as comparison of final models with initial and medial models.

CONCLUSION

The final consonant subsystem of the Tajik language comprises 125 structural types in the working sample. Its core consists of 114 binary sequences (91.2 %); ten ternary and one quaternary sequence constitute the periphery, predominantly associated with international terminology and onomastics. Within the binary inventory, the III+III and C+III macro-models dominate (40 types each), while in 82 of the 114 types the final component is a stop or fricative obstruent, confirming the general tendency of word-final position toward decreasing sonority.

The phonetic organization does not mechanically correspond to the orthographic representation: positional devoicing of final voiced obstruents, place assimilation of /n/, and fricative realization of /κ/ reduce articulatory complexity and produce the neutralization of orthographically distinct clusters. The glottal stop and final geminates form special classes whose definitive interpretation requires experimental phonetic verification. The sonority principle adequately describes the general tendency, but the stable types -бp, -дp, -xp, -шш demonstrate that historical integration and lexical entrenchment can license a local rise in sonority. Therefore, the model of Tajik word-final structure should be regarded as gradual and stratified.

The proposed classification provides a basis for comparative research, for the improvement of pronunciation dictionaries, and for the development of rules for the automatic conversion of orthographic forms into speech. The next stage of the research should combine the typological table with corpus frequency data and acoustic evidence.

REFERENCES

1. Avanesov R. I. 1984. *Russian Literary Pronunciation*. 6th ed. Moscow: Prosveshchenie.
2. Bondarko, L. V. 1977. *The Sound Structure of Modern Russian*. Moscow: Prosveshchenie.
3. Rules of Tajik orthography and punctuation marks. 2021. Resolution No. 268 of the Government of the Republic of Tajikistan. 30 Jun 2021. Dushanbe.
4. Rastorgueva, V. S. 1954. *An Outline of Tajik Grammar*. Issue 1: Phonetics and Morphology. Moscow: Publishing House of the Academy of Sciences of the USSR.
5. Rahmatullozoda, S. R. (Ed.) 2025. *Tajik Orthographic Dictionary*. Dushanbe: Mir Poligrafii.
6. Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In J. Kingston, M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 283-333). Cambridge: Cambridge University Press.
7. Goldsmith, J. A. 1990. *Autosegmental and Metrical Phonology*. Oxford: Basil Blackwell.
8. Greenberg, J. H. 1978. Some generalizations concerning initial and final consonant clusters. In J. H. Greenberg, C. A. Ferguson, E. A. Moravcsik (Eds.), *Universals of Human Language* (pp. 243-279). Vol. 2: Phonology. Stanford: Stanford University Press.
9. Kager, R. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.
10. Ladefoged, P. & Maddieson, I. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.
11. McCarthy, J. J. 1986. OCP effects: Gemination and antigemination. *Linguistic Inquiry*, 17/2, 207-263.
12. Perry, J. R. 2005. *A Tajik Persian Reference Grammar*. Leiden; Boston: Brill.

13. Selkirk E. O. 1982. The syllable. In H. van der Hulst & N. Smith (Eds.), *The Structure of Phonological Representations* (pp. 337-383). Part II. Dordrecht: Foris.
14. *Tajik Hunspell Dictionary (tg_TG.dic)*: machine-readable dictionary database containing 68 665 entries. Version 1.0.0. Apache License 2.0.
15. Vennemann, T. 1988. *Preference Laws for Syllable Structure and the Explanation of Sound Change*. Berlin; New York; Amsterdam: Mouton de Gruyter.

PANJIYEV NORMAXMAT

A DOCENT, DSC,

TERMEZ STATE UNIVERSITY, UZBEKISTAN.

E-MAIL: <PANJIEV.RUSLIT@GMAIL.COM>