

Кандидат филологических наук, доцент
Термезского государственного университета
E-mail: panjiev.ruslit@gmail.com
ORCID: 0000-0002-7594-1011

КОНСОНАНТНЫЕ КЛАСТЕРЫ УЗБЕКСКОГО ЯЗЫКА: КРИТЕРИИ РАЗГРАНИЧЕНИЯ ГРАФИКИ И ФОНОЛОГИИ

Аннотация. Статья посвящена фонологической верификации консонантных кластеров современного узбекского языка. Цель исследования – выработать критерии, позволяющие отличать подлинный кластер от графической последовательности согласных, аффрикаты, диграфа, геминаты и произносительного варианта. Материал составили нормативные орфографический, толковый и орфоэпический словари; применены фонемный, дистрибутивный, морфемный и лексикографический методы. Предлагается модель G–Ф–R, разграничивающая графическую форму, фонемную структуру и нормативную реализацию. Она позволяет корректно интерпретировать *sh*, *ch*, *ng*, удвоенные согласные, сочетания *C+u* и заимствования с энпентезой. Научный результат – воспроизводимая процедура отбора, обеспечивающая достоверность последующего количественного и сопоставительного анализа.

Ключевые слова: узбекский язык; консонантный кластер; фонемная сегментация; графика; орфоэпия; аффриката; гемината; энпентеза; фонотактика.

O'ZBEK TILINING KONSONANT KLASTERLARI: GRAFIKA VA FONOLOGIYANI FARQLASH MEZONLARI

Annotatsiya. Maqolada hozirgi o'zbek tilidagi undosh klasterlarni fonologik tekshirish metodikasi yoritiladi. Tadqiqotning maqsadi haqiqiy klasterlarni undosh harflar ketma-ketligi, digraf, affrikata, geminata va talaffuz variantidan ajratadigan amaliy mezonlarni belgilashdir. Tadqiqot materiallari imlo, izohli va orfoepik lug'atlardan olindi; fonemik, distributiv, morfemik hamda leksikografik tahlil usullari qo'llandi. Grafik yozuv G, fonemik tuzilish Ф va nutqdagi me'yoriy

realizatsiyadan iborat G–Φ–R modeli taklif qilinadi. U sh, ch, ng, qo 'sh undoshlar, C+y ketma-ketliklari va epentezali o'zlashmalarni farqlashga imkon beradi. Natijada klasterlarni tanlashning takrorlanuvchi tartibi ishlab chiqildi; u miqdoriy va qiyosiy tahlilning ishonchligini oshiradi.

Kalit so'zlar: o'zbek tili; undoshlar klasteri; fonemik segmentatsiya; grafika; orfoepiya; affrikata; geminata; epenteza; fonotaktika.

CONSONANT CLUSTERS IN THE UZBEK LANGUAGE: CRITERIA FOR DISTINGUISHING ORTHOGRAPHY FROM PHONOLOGY

Abstract. The article develops a procedure for the phonological validation of consonant clusters in Modern Uzbek. It aims to distinguish a genuine cluster from a graphic sequence of consonant letters, a digraph, an affricate, a geminate, or a pronunciation variant. The material comprises normative orthographic, explanatory, and orthoepic dictionaries; the analysis combines phonemic, distributional, morphemic, and lexicographic methods. The proposed G–Φ–R model differentiates graphic form, phonological structure, and normative realization. It clarifies the status of sh, ch, ng, doubled consonants, C+y sequences, and loanwords with epenthesis. The main result is a reproducible selection procedure that improves the reliability of subsequent quantitative and comparative analysis.

Keywords: uzbek language; consonant cluster; phonemic segmentation; orthography; orthoepy; affricate; geminate; epenthesis; phonotactics.

Исследование консонантных кластеров в узбекском языке нередко начинается с поиска соседних согласных букв в словарном реестре. Такой приём полезен как технический первый шаг, однако он не может быть принят за конечную процедуру. Графическая последовательность ещё не доказывает наличие двух или более самостоятельных согласных фонем. В латинской графике узбекского языка одни и те же внешние признаки создают принципиально разные фонологические ситуации: сочетания sh и ch передают единые фонемы, ng в зависимости от лексемы и морфемного членения требует

отдельного решения, а удвоенная буква способна обозначать геминатную реализацию. Кроме того, в заимствованной лексике написание может сохранять сложный согласный комплекс, тогда как в речевом потоке он варьирует из-за эпентезы, протезы или упрощения.

Поэтому основной вопрос состоит не в том, сколько согласных букв стоит рядом, а в том, какие сегменты образуют нормативную фонемную структуру слова и как они реализуются в речи. Без такой проверки количественные показатели становятся несопоставимыми: в один массив попадают фонемные кластеры, аффрикаты, графические диграфы, геминаты и отдельные варианты произношения. Для узбекской фонотактики это особенно существенно, поскольку исконная слоговая организация взаимодействует с богатым слоем давно освоенных и новейших заимствований, а также с вариативностью графического оформления [1, 3-с].

Цель статьи – разработать систему критериев фонологической верификации консонантных кластеров современного узбекского языка. Для её достижения решаются следующие задачи: уточняется рабочее определение кластера; разграничиваются графический, фонемный и орфоэпический уровни описания; предлагается последовательность диагностических операций; анализируются пограничные случаи, наиболее опасные для автоматического и ручного подсчёта. Статья не повторяет инвентаризацию и статистическое описание начальных или конечных сочетаний: её предметом является методика принятия решения о том, является ли конкретная последовательность собственно кластером.

В фонологическом смысле консонантным кластером следует считать непрерывную последовательность не менее двух самостоятельных согласных фонем, расположенную в определённой позиции словоформы и не разделённую гласным ядром. Для абсолютного начала слова это максимальная цепь согласных от первого сегмента до первой гласной; для конца слова – максимальная цепь после последней гласной до правой границы слова.

Внутренняя позиция может анализироваться по тому же принципу, но в настоящей статье она привлекается лишь тогда, когда необходима морфемная проверка спорной записи. Такое понимание соотносится с принципами слоговой организации и сонорности в описании последовательностей согласных [7].

Такое определение опирается на фонологическое различие единиц, а не на орфографическое оформление. Фонема устанавливается по совокупности её различительных функций и дистрибуции, тогда как графема является средством письма и может передавать одну единицу несколькими знаками либо один знак использовать в неодинаковых контекстах [5]. В описаниях узбекской фонетики данное различие особенно важно для сложных графем и периферийной заимствованной лексики [1, 4-с].

Рабочее определение включает три обязательных условия. Во-первых, каждый компонент должен быть фонологически самостоятельным: если два знака представляют одну аффрикату или один носовой сегмент, двучленного кластера нет. Во-вторых, между компонентами не должно быть фонемной гласной; факультативный вставной гласный в реальной речи фиксируется отдельно. В-третьих, последовательность должна быть максимальной в своей позиции: в слове с трёхкомпонентным началом нельзя учитывать одновременно полный комплекс и произвольно выделенную из него двучленную часть. Последнее правило устраняет искусственное разрастание перечня типов.

Следовательно, автоматический поиск по буквенным шаблонам даёт не итоговый корпус, а список кандидатов. Сам термин «кандидат» важен методологически: запись включается в строгий фонемный корпус только после того, как установлены границы сегментов, их количество и позиционный статус. Такой порядок согласуется с общим принципом разграничения фонетической реализации и фонологической структуры [6].

Для описания проблемных случаев предлагается трёхуровневая модель G–Ф–R. Символ G обозначает графическую форму леммы после технической нормализации; Ф – её фонемную запись; R – нормативно предполагаемую либо наблюдаемую речевую реализацию. Переход $G \rightarrow \Phi$ не является механическим: именно на этом этапе проверяются многознаковые графемы, аффрикаты, морфемные границы и геминатные случаи. Переход $\Phi \rightarrow R$ отражает позиционное оглушение, ассимиляцию, эпентезу, протезу, упрощение и другие процессы, которые могут менять звучание, не создавая новой словарной модели.

Рассмотрим принцип действия модели на простых примерах. Графическая начальная последовательность shl- в слове shlang состоит из трёх букв, но при фонемной сегментации даёт /ʃl/: графема sh передаёт один щелевой сегмент, а l-второй. Напротив, ch в слове chiroq передаёт аффрикату /tʃ/ и не должно быть включено в перечень двучленных кластеров. При этом буквенная длина комплекса shl-больше, чем ch, но фонемная компонентность соотносится с ними обратным образом. Именно такие случаи делают буквенный подсчёт ненадёжным.

Уровень R необходим, чтобы не смешивать структуру слова и произносительную вариативность. В заимствовании типа plan фонемная последовательность /pl/ может сохраняться в литературной реализации, однако в части речевых ситуаций возможно появление эпентетического гласного. В таком случае основной тип остаётся фиксируемым как /pl/, а вариант получает отдельную метку эпентезы. Иначе одна и та же лексема будет ошибочно превращаться в несколько разных структурных единиц. Аналогично, в финали слова labz буквенная группа – bz в нормативной предпаузальной реализации может давать [ps]: качество обоих компонентов меняется, но их число не становится равным одному.

Узбекская орфография требует предварительной технической нормализации. Апострофы и кавычкообразные знаки в o' и g' могут быть

представлены различными компьютерными символами, что порождает формальные дубликаты одной леммы. Нормализация не подменяет лингвистический анализ, но обеспечивает одинаковую исходную запись для него. Только после унификации регистра, апострофов, дефисов и служебных помет допустимо применять регулярные поисковые шаблоны к словарному материалу [8, 9-с].

Диагностическая процедура и система статусов

Предлагаемая процедура состоит из шести связанных операций. Первая операция – выделение леммы и её графическая нормализация. Вторая – распознавание многознаковых графем и буквенных сочетаний, которые потенциально не равны двум фонемам. Третья – фонемная сегментация с опорой на описания фонологической системы и орфоэпический словарь [9, 10-b]. Четвёртая – определение максимальной непрерывной цепи согласных в нужной позиции. Пятая – проверка морфемной границы, если она влияет на интерпретацию *ng*, удвоения либо вариативного гласного. Шестая – фиксация произносительного варианта как отдельного свойства записи, а не как самостоятельного кластера.

Результат проверки следует оформлять не словесной пометой «есть стечение» или «нет стечения», а компактным кодом. Для собственно кластера указываются позиция и число сегментов, например CL-INIT-2 или CL-FIN-3. Для аффрикаты используется статус AFF-1, для однофонемной реализации *ng* – NAS-1, для геминатного случая – GEM, для непринятого до проверки решения – PEND. Если речевая реализация содержит эпентезу, добавляется VAR-EPI; для конечного оглушения уместна метка VAR-DEV. Такая разметка позволяет хранить в единой базе все наблюдения, не смешивая их при подсчёте.

Главное преимущество этой модели заключается в воспроизводимости. Другой исследователь может увидеть не только итоговый статус единицы, но и основание решения: какая графическая форма была исходной, какая

фонемная сегментация принята и какой орфоэпический вариант отмечен. В условиях работы с разными словарями и электронными реестрами это предотвращает скрытую замену одной единицы анализа другой. Лексикографические источники выполняют при этом разные функции: орфографический словарь задаёт нормативный реестр, толковый подтверждает самостоятельность и значение леммы, орфоэпический уточняет произносительный статус [10, 9-с].

Таблица 1 – Диагностические статусы графических последовательностей согласных

Графический кандидат	Фонемное решение	Статус	Комментарий
shl- в shlang	/ʃl/	CL-INIT-2	Две фонемы при трёх буквах: sh передаёт один целевой согласный.
ch в chiroq	/tʃ/	AFF-1	Аффриката не расчленяется на два самостоятельных сегмента.
ng в конкретной лемме	/ŋ/ или /n+g/	NAS-1 / CL	Решение зависит от фонемного состава и морфемной границы; буквенная форма недостаточна.
pl- в plan	/pl/	CL-INIT-2	При эпентезе в речевой реализации фиксируется дополнительная метка VAR-EPI, а не новый тип.
-bz в labz	[ps]	CL-FIN-2	Позиционное оглушение меняет качество компонентов, но не устраняет их двусегментность.
Удвоенная согласная	[C:] / C+C	GEM	Статус устанавливается отдельно; двойное написание не даёт права автоматически считать кластер.
C+y в заимствовании	/Cj/ или иная реализация	PEND	Требует индивидуальной орфоэпической и фонологической проверки.

Примечание. CL – собственно кластер; AFF – аффриката; GEM – геминатный случай; PEND – решение отложено до специальной проверки; VAR-EPI – эпентетический вариант.

Диграфы, аффрикаты и последовательность ng

Первую группу составляют записи, в которых два или более буквенных знака не равны числу согласных фонем. Графемы sh и ch должны обрабатываться как целые единицы ещё на этапе подготовки кандидатов. В противном случае системы автоматического поиска будут ошибочно создавать ложные сочетания, например выделять sh как два согласных. Фонемная интерпретация аффрикаты также не должна зависеть от её графической разложимости: наличие в артикуляции смычной и щелевой фаз не превращает /tʃ/ в последовательность /t+j/ для целей кластерного подсчёта.

Последовательность ng требует более осторожного подхода. В части слов она соотносится с одной заднеязычной носовой фонемой /ŋ/ и должна получать статус NAS-1. В других контекстах, в особенности при наличии убедимой морфемной границы или при иной фонемной трактовке конкретной лексемы, возможна последовательность /n+g/. Поэтому правило «ng всегда кластер» так же ошибочно, как и правило «ng никогда не кластер». Практически это означает обязательное сопоставление записи с лексикографическим толкованием, морфемной структурой и орфоэпическими сведениями. Только такое тройное подтверждение позволяет снять неоднозначность.

Разграничение имеет и количественное следствие. Если не отделить /ŋ/ от /n+g/, число финальных и медиальных двухсегментных последовательностей будет завышено, а распределение носовых компонентов окажется искажённым. Более того, в сравнительном исследовании разные языки могут использовать сходное буквенное сочетание при различном фонемном членении; перенос графического решения из одного языка в другой недопустим.

Удвоенные согласные и геминатный статус

Вторую группу образуют удвоенные согласные. В графике удвоение выглядит как два одинаковых знака, однако фонологически оно может интерпретироваться как один длительный согласный, как последовательность двух одинаковых фонем либо как орфографическое наследие, не равное полноценной сегментной длительности. Эти возможности нельзя свести к единому автоматическому правилу. Поэтому двойная буква должна сначала быть выделена в отдельный класс GEM и лишь после специального фонологического обоснования включаться или не включаться в строгий кластерный корпус.

Такое решение не означает, что геминаты не важны для фонотактики. Напротив, они требуют самостоятельного описания, поскольку длительность, морфемная граница и орфографическая норма могут по-разному участвовать в их формировании. Но отождествлять геминату с гетероорганном или гетеросегментным кластером методологически неверно. В частности, статистика типов C_1C_2 должна быть отделена от статистики $C:$, иначе один и тот же признак – удвоение буквы – будет измерять две разные лингвистические величины.

Последовательности C+у и роль орфоэпической проверки

Особую периферию составляют сочетания вида C+у в заимствованных словах. Буква у может передавать самостоятельный аппроксимант /j/, участвовать в передаче качества предшествующего согласного или реализовываться вариативно в зависимости от степени освоенности слова. Следовательно, такие записи не следует заранее присоединять к двумсегментным кластерам и столь же преждевременно исключать из них. Оптимальным решением является статус PEND с обязательной проверкой по орфоэпическому словарю и, при необходимости, по материалам реальной речи. В этом пункте особенно заметна разница между фонологической системой узбекского языка и русским описанием мягкости. Для узбекского

литературного языка противопоставление твёрдых и мягких согласных не образует той системной корреляции, которая характерна для русского. Поэтому графическую форму $C+u$ нельзя автоматически объяснять как «мягкий согласный плюс несуществующий второй компонент» и нельзя без проверки считать парой $C+j$. Решение должно следовать из сегментации конкретной леммы, а не из внешнего сходства с русской графикой [1, 4-b].

Заемствованные кластеры, эпентеза и конечное оглушение

Заемствования демонстрируют, почему необходимо различать Φ и R . Сложная инициальная последовательность может быть нормативно закреплена в лексеме, но в спонтанной речи под влиянием привычной слоговой структуры расчленяться вставным гласным. Эпентеза не отменяет фонемного анализа исходного слова, но сообщает о степени адаптационной вариативности. В записи базы данных это лучше передавать сочетанием основного кода кластера и дополнительной метки VAR-EPI, а не создавать новый «кластерный тип» с гласным между компонентами. Общие закономерности вставки гласных в заимствованиях хорошо известны в типологии заимствований, однако конкретная реализация должна описываться по материалу языка [11, 12-b].

В финали действует противоположный по характеру процесс: два фонемных компонента могут сохранять непрерывность, но подвергаться позиционному оглушению или регрессивному согласованию. Запись $-bz$ в конце слова не следует трактовать как неизменную фонетическую пару $[b+z]$, если нормативная реализация отражает иные признаки звонкости. Однако и здесь изменение качества не должно смешиваться с изменением количества: $[ps]$ остаётся двухсегментной реализацией. Разделение фонемного и орфоэпического слоёв делает видимыми оба факта - сохранение структуры и изменение её звучания.

Выводы и рекомендации

Предложенная процедура полезна прежде всего для создания проверяемой словарной базы. Если каждую единицу снабжать полями G, Ф, R, позиция, морфемная граница, структурный код и источник проверки, можно выполнять автоматический поиск без утраты лингвистического контроля. Машина быстро формирует перечень кандидатов, а специалист принимает решение в тех точках, где графика неоднозначна. Такой подход особенно продуктивен при работе с крупным орфографическим реестром, содержащим общеупотребительную, терминологическую и заимствованную лексику.

Для количественного анализа методика устраняет несколько типичных ошибок: двойной учёт длинного кластера и его частей, включение аффрикат в число двучленных последовательностей, механическое превращение ng в /n+g/, смешение геминат с разнородными сочетаниями и подмена фонемной структуры одним речевым вариантом. Благодаря этому показатели типа «число кластеров», «число лемм» и «число структурных моделей» получают ясное различие. Они могут быть сопоставлены между исследованиями только при совпадении единицы анализа и правил фильтрации.

Методика имеет и дидактическое значение. В преподавании узбекской фонетики учащимся часто приходится объяснять, почему число букв и число звуков не совпадают и почему заимствованное слово может иметь несколько произносительных вариантов. Трёхуровневая схема позволяет наглядно показать последовательность: сначала написание, затем фонемный состав, затем реальная реализация. Это снижает риск орфографического подхода к фонетическим задачам и помогает строить корректные сопоставления с русским и другими языками.

Наконец, разграничение статусов создаёт основу для дальнейшего исследования не только позиционной дистрибуции, но и фонологической адаптации заимствований, слоговой организации, орфоэпических норм и вариативности. При этом новый материал не будет искусственно смешан с уже известными структурными типами: каждый новый пример сначала проходит

одну и ту же процедуру верификации, а уже потом попадает в тематический анализ. Сформулированная методика развивает ранее выполненные структурно-статистические наблюдения и переводит их в единый проверяемый формат [13].

Основные выводы

1. Консонантный кластер в современном узбекском языке должен определяться как непрерывная последовательность самостоятельных согласных фонем, а не как ряд соседних согласных букв.

2. Для надёжного анализа необходима трёхуровневая модель $G - \Phi - R$, различающая графическую запись, фонемную структуру и нормативную речевую реализацию.

3. Графемы *sh* и *ch*, а также контекстно неоднозначное *ng*, удвоенные согласные и последовательности *C+u* требуют специальных статусов и не могут автоматически включаться в двучленный кластерный корпус.

4. Эпентеза, протеза, конечное оглушение и ассимиляция должны фиксироваться как свойства реализации, а не как основание для замены фонемной модели лексемы.

5. Процедура графической нормализации, фонемной сегментации, позиционного и морфемного контроля обеспечивает воспроизводимость отбора и предотвращает завышение либо смещение количественных показателей.

6. Методика является самостоятельным этапом исследования: она формирует корректный материал для последующих работ по фонотактике, орфоэпии и сопоставительной типологии узбекского языка.

Использованная литература

1. Абдуазизов А. Ўзбек тили фонологияси ва морфонологияси. – Тошкент: Ўқитувчи, 1992. – 138 б.

2. Кононов А. Грамматика современного узбекского литературного языка. – Москва; Ленинград: Издательство Академии наук СССР, 1960. – 446 с.
3. Jamolxonov H. *Hozirgi o'zbek adabiy tili*. – Toshkent: Talqin, 2005. – 272 b.
4. Sayfullayeva R., Mengliyev B., Raipova L., Qurbonova M., Abuzalova M., Yo'ldosheva D. *Hozirgi o'zbek tili*. – Toshkent: Innovatsiya-Ziyo, 2020. – 552 b.
5. Трубецкой Н. Основы фонологии / перевод с немецкого А.А.Холодовича; под редакцией С.Д.Кацнельсона. – Москва: Издательство иностранной литературы, 1960. – 372 с.
6. International Phonetic Association. *Handbook of the International Phonetic Association*. – Cambridge: Cambridge University Press, 1999. – 204 p.
7. Clements G. *The Role of the Sonority Cycle in Core Syllabification // Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech / ed. by J.Kingston, M.Beckman*. – Cambridge: Cambridge University Press, 1990. – P. 283–333.
8. Mahmudov N., Madvaliyev A., Begmatov E. *O'zbek tilining imlo lug'ati: 85000 dan ortiq so'z*. – Toshkent: Akademnashr, 2013. – 526 b.
9. *O'zbek tilining izohli lug'ati: 6 jildli / mas'ul muharrirlar A.Madvaliyev, D.Xudayberganova*. – Toshkent: G'afur G'ulom nomidagi nashriyot-matbaa ijodiy uyi, 2022–2023. 6 jild.
10. Mamatov J., Chiniqulov N. *O'zbek tilining orfoepik lug'ati*. – Toshkent: Donishmand ziyosi, 2021. – 112 b.
11. Kang Y. *Loanword Phonology // The Blackwell Companion to Phonology / edited by M. van Oostendorp, C.J.Ewen, E. Hume, K.Rice*. – Oxford: Wiley-Blackwell, 2011. – P. 1003–1026.
12. Hall N. *Cross-linguistic Patterns of Vowel Intrusion // Phonology*. – 2006. Vol. 23, no. 3. – P. 387–429.
13. Панжиев Н. Исследование структурно-статистических характеристик консонантных кластеров в узбекском языке // *Educational Research in Universal Sciences*. – 2023. Т. 2, № 7. – С. 38–46.