УДК 519.85:81'322

α-СПРАВЕДЛИВАЯ ОПТИМИЗАЦИОННАЯ МОДЕЛЬ И ЧИСЛЕННЫЕ АЛГОРИТМЫ ВЫБОРА ОБЩЕЙ ЛЕКСИЧЕСКОЙ НОРМЫ ТЮРКСКИХ ЯЗЫКОВ

[1*]***Бакаев И.И.,*** [2]***Ахмеджанова Д.А.***

[*]i.i.bakaev@buxdu.uz, [2]daxmadjonova@gmail.com

[1]Бухарский государственный университет, 200114, Узбекистан, г. Бухара, ул. М. Икбола, 11;

[2]Научно-исследовательский институт развития цифровых технологий и искусственного интеллекта, 100125, Узбекистан, г. Ташкент, м-в Буз-2, 17А

Задача выбора общей лексической нормы тюркских языков поставлена как задача дискретной оптимизации, обеспечивающая справедливую понятность для носителей разных языков. Для каждого понятия из множества форм-кандидатов выбирается одна форма; понятность для носителя каждого языка оценивается посредством взвешенного редакционного расстояния на основе фонетических признаков, а выбор определяется α-справедливой целевой функцией и ограничениями фонологической согласованности. Доказано, что модель взвешенного усреднения является частным случаем при $\alpha = 0$, а критерий max-min — пределом при $\alpha \to \infty$. Разработаны точная формулировка задачи в виде задачи смешанного целочисленного линейного программирования и приближённый алгоритм на основе релаксации, зеркального подъёма, округления и локального поиска. Вычислительный эксперимент на 1016 понятиях пяти тюркских языков из базы NorthEuraLex показал, что если при взвешенном усреднении понятность для наименее обеспеченного языка (казахского) равна 0,528, то в max-min решении она возрастает до 0,704 (на 33,2 %), тогда как взвешенное по численности населения среднее снижается лишь на 6,0 %. Предложенный приближённый алгоритм отличается от точного решения менее чем на 0,03 % и при $\alpha = 1$ работает в 49 раз быстрее.

Ключевые слова: тюркские языки, общая норма, α-справедливость, max-min оптимизация, смешанное целочисленное программирование, зеркальный подъём, фонетическое расстояние, NorthEuraLex

Цитирование: *Бакаев И.И., Ахмеджанова Д.А.* α-справедливая оптимизационная модель и численные алгоритмы выбора общей лексической нормы тюркских языков // Проблемы вычислительной и прикладной математики. – 2026. – №5(76). – С. 137-147.

DOI: 10.71310/pcam.5_76.2026.09

1 Введение

Идея создания общей нормы для родственных языков имеет в тюркологии давнюю историю. В концепции «усреднённого тюркского языка» Б. Каримова и Ш. Муталова [1] на каждом уровне языка выделяется общий фонд, единицы которого предлагаются в качестве нормы «усреднённого» языка. Недавно эта идея получила математическое оформление: в работе [2] формы каждой языковой единицы получают взвешенные оценки по критериям частотности, распространённости, простоты и культурной релевантности, и выбирается форма с наибольшим баллом.

Однако выбор, основанный на взвешенном усреднении, обладает существенным недостатком: формы языков с большим числом носителей систематически доминируют, вследствие чего норма оказывается менее понятной для носителей малых языков. Это явление хорошо известно в теории распределения: максимизация суммарной полезности (утилитарный критерий) не защищает наименее обеспеченного участника [3, 4]. Семейство α-справедливых функций, предложенное Мо и Уолрендом [5] для распределения ресурсов в компьютерных сетях, с помощью одного параметра обеспечивает непрерывный переход между утилитарным критерием ($\alpha = 0$), пропорциональной справедливостью ($\alpha = 1$, решение Нэша [6]) и max-min справедливостью ($\alpha \to \infty$).

Показано, что фонетическое расстояние является одним из основных факторов при измерении взаимопонятности родственных языков [7, 8]. Расстояние Левенштейна и его варианты, взвешенные по фонологическим признакам [8–10], широко применяются для оценки различий в произношении между диалектами и родственными языками.

Целью настоящей статьи является постановка задачи выбора общей лексической нормы тюркских языков как задачи α-справедливой дискретной оптимизации, исследование её свойств и разработка эффективных численных алгоритмов её решения. Основные результаты статьи состоят в следующем:

1. предложена модель α-справедливого выбора нормы с ограничениями фонологической согласованности; доказано, что модель взвешенного усреднения из [2] является её частным случаем при $\alpha = 0$, а критерий max-min — её пределом при $\alpha \to \infty$;
2. дана формулировка задачи в виде задачи смешанного целочисленного линейного программирования (MILP) и апостериорная оценка точности на основе релаксации;
3. разработан приближённый алгоритм на основе зеркального подъёма, округления и локального поиска;
4. проведён вычислительный эксперимент на 1016 понятиях пяти тюркских языков из базы NorthEuraLex [11] и дана количественная оценка «цены справедливости».

2 Постановка задачи

2.1 Обозначения и показатель понятности

Пусть $a = 1, \ldots, A$ — языки, $n = 1, \ldots, N$ — понятия (смысловые единицы), C_n — множество форм-кандидатов для понятия n. В качестве кандидатов берутся формы, существующие в языках: $C_n = \bigcup_a Y_{a,n}$, где $Y_{a,n}$ — множество форм, выражающих понятие n в языке a. Каждая форма рассматривается как последовательность фонем (символов IPA).

Понятность формы $c \in C_n$ для носителя языка a оценивается следующим образом:

$$p_{a,n,c} = \max_{y \in Y_{a,n}} \left(1 - \frac{d(c,y)}{\max\left(|c|, |y|\right)}\right), \tag{1}$$

где $|c|$ — длина формы, $d(\cdot,\cdot)$ — взвешенное редакционное расстояние, вычисляемое методом динамического программирования [9]:

$$D_{i,j} = \min\left\{D_{i-1,j} + 1,\ D_{i,j-1} + 1,\ D_{i-1,j-1} + \varkappa\left(s_i, t_j\right)\right\}, \qquad D_{i,0} = i,\ D_{0,j} = j. \tag{2}$$

Стоимость замены $\varkappa(u,v)$ определяется через фонологические признаки: $\varkappa = 0$, если $u = v$; $\varkappa = 0{,}1$, если фонемы различаются только долготой; $\varkappa = 0{,}15 + 0{,}6\,\delta_V(u,v)$

для гласных, $\varkappa = 0{,}15 + 0{,}6\,\delta_C(u,v)$ для согласных; $\varkappa = 1$ для пары «гласный — согласный». Здесь

$$\delta_V = \frac{1}{3}\left(\frac{|\Delta h|}{3} + \frac{|\Delta q|}{2} + |\Delta l|\right), \qquad \delta_C = \frac{1}{3}\left(\frac{|\Delta o|}{7} + [\,\text{способ различен}\,] + |\Delta j|\right), \quad (3)$$

h — подъём гласного, q — ряд, l — огубленность; o — место образования согласного, j — звонкость; $[\cdot]$ — скобка Айверсона.

Пусть $x_{n,c} \in \{0,1\}$ — переменная, указывающая, выбрана ли форма c для понятия n. Понятность нормы для носителя языка a равна

$$U_a(x) = \sum_{n=1}^{N} f_n \sum_{c \in C_n} x_{n,c}\, p_{a,n,c}, \qquad f_n \geqslant 0, \quad \sum_{n=1}^{N} f_n = 1, \quad (4)$$

где f_n — относительная частота понятия в тексте.

2.2 α-справедливая целевая функция

Распределение понятности между языками оценивается с помощью α-справедливой функции [5]:

$$F_\alpha(x) = \sum_{a=1}^{A} w_a\, \varphi_\alpha\left(U_a(x)\right), \qquad \varphi_\alpha(u) = \begin{cases} \dfrac{u^{1-\alpha}}{1-\alpha}, & \alpha \geqslant 0,\ \alpha \neq 1, \\ \ln u, & \alpha = 1, \end{cases} \quad (5)$$

где $w_a > 0$, $\sum_a w_a = 1$ — веса языков (например, зависящие от числа носителей). При $\alpha \to \infty$ в качестве целевой функции принимается $F_\infty(x) = \min\limits_a U_a(x)$.

2.3 Ограничения фонологической согласованности

В тюркских языках существуют регулярные ряды соответствий начальных согласных: например, *gör-*, *dil* в огузских языках и *kör-*, *til* в кыпчакских и карлукских языках. Если в норме для одного слова выбран звонкий вариант, а для другого — глухой, норма утрачивает внутреннюю согласованность. Для учёта этого вводятся классы соответствий $k = 1, \ldots, K$ и множества альтернативных правил R_k в каждом классе; $z_r \in \{0,1\}$ — индикатор выбора правила r. Через $\mathcal{D}$ обозначено множество кандидатов, зависящих от правил, через $\rho(n,c)$ — правило, от которого зависит кандидат (n,c).

В результате получаем следующую задачу:

$$(P_\alpha): \qquad \max_{x,\,z} \ F_\alpha(x) \quad (6)$$

$$\sum_{c \in C_n} x_{n,c} = 1, \quad n = 1, \ldots, N, \quad (7)$$

$$x_{n,c} \leqslant z_{\rho(n,c)}, \quad (n,c) \in \mathcal{D}, \quad (8)$$

$$\sum_{r \in R_k} z_r = 1, \quad k = 1, \ldots, K, \quad (9)$$

$$x_{n,c} \in \{0,1\}, \quad z_r \in \{0,1\}. \quad (10)$$

Ограничение (7) обеспечивает выбор ровно одной формы для каждого понятия, а ограничения (8)–(9) — выбор ровно одного правила в каждом классе соответствий и исключение из нормы противоречащих ему форм.

3 Свойства модели

Утверждение 1 (случай $\alpha = 0$)**.** *Если $\alpha = 0$ и вектор правил z фиксирован, то задача (P_0) распадается по понятиям, и её решение имеет вид*

$$c_n^* = \arg\max_{c \in C_n(z)} \sum_{a=1}^{A} w_a \, p_{a,n,c}, \qquad n = 1, \ldots, N, \tag{11}$$

где $C_n(z)$ — кандидаты, не противоречащие ограничению (8)*. Глобальное решение находится перебором $\prod_k |R_k|$ комбинаций правил за $O\big(\prod_k |R_k| \cdot A \sum_n |C_n|\big)$ операций. При $\mathcal{D} = \varnothing$* (11) *по существу совпадает с критерием из [2] — взвешенным оцениванием и выбором формы с наибольшим баллом.*

Доказательство. Согласно (4), $F_0(x) = \sum_a w_a U_a(x) = \sum_n f_n \sum_c x_{n,c} \sum_a w_a p_{a,n,c}$, т. е. целевая функция линейна и распадается на слагаемые по понятиям. При фиксированном z ограничение (8) лишь исключает некоторых кандидатов, а (7) является отдельным ограничением для каждого понятия. Следовательно, каждое слагаемое можно максимизировать независимо.

Утверждение 2 (релаксация и оценка точности)**.** *Пусть z фиксирован и $\alpha \geqslant 0$. Задача $(\bar{P}_\alpha(z))$, полученная заменой условия $x_{n,c} \in \{0, 1\}$ в* (10) *на $x_{n,c} \in [0, 1]$, является задачей максимизации вогнутой функции на выпуклом множестве. Для её оптимального значения $\bar{F}_\alpha(z)$ и любого допустимого целочисленного решения $\widehat{x}$ справедлива оценка*

$$F_\alpha^* - F_\alpha(\widehat{x}) \leqslant \max_z \bar{F}_\alpha(z) - F_\alpha(\widehat{x}). \tag{12}$$

При $\alpha = \infty$ задача $(\bar{P}_\infty(z))$ является задачей линейного программирования.

Доказательство. $U_a(x)$ — линейная функция от x. При $u > 0$ имеем $\varphi'_\alpha(u) = u^{-\alpha} > 0$ и $\varphi''_\alpha(u) = -\alpha u^{-\alpha-1} \leqslant 0$, т. е. φ_α возрастает и вогнута. Композиция вогнутой функции с линейной вогнута, сумма с неотрицательными коэффициентами также вогнута. Минимум линейных функций вогнут и через гипограф выражается линейными ограничениями. Допустимое множество, будучи произведением симплексов, выпукло и компактно. Целочисленные допустимые точки принадлежат релаксации, поэтому $F_\alpha^* \leqslant \max_z \bar{F}_\alpha(z)$, откуда следует (12).

Утверждение 3 (предел при $\alpha \to \infty$)**.** *Пусть множество допустимых решений X конечно и $w_a > 0$, $U_a(x) > 0$ для всех a и $x \in X$. Тогда существует такое α_0, что при $\alpha > \alpha_0$ каждое решение задачи (P_α) максимизирует также и $\min_a U_a(x)$.*

Доказательство. При $\alpha > 1$ максимизация F_α равносильна минимизации $G_\beta(x) = \sum_a w_a U_a(x)^{-\beta}$, $\beta = \alpha - 1 > 0$. Положим $m(x) = \min_a U_a(x)$, $\underline{w} = \min_a w_a$, $\overline{w} = \max_a w_a$. Тогда

$$\underline{w}\, m(x)^{-\beta} \leqslant G_\beta(x) \leqslant A\, \overline{w}\, m(x)^{-\beta}. \tag{13}$$

Пусть x^* — max-min решение, а $y \in X$ таково, что $m(y) < m(x^*)$. Согласно (13),

$$\frac{G_\beta(x^*)}{G_\beta(y)} \leqslant \frac{A\,\overline{w}}{\underline{w}} \left(\frac{m(y)}{m(x^*)}\right)^{\beta} \xrightarrow[\beta \to \infty]{} 0,$$

следовательно, начиная с некоторого β_y, выполняется $G_\beta(x^*) < G_\beta(y)$ и y не является оптимальным. Поскольку X конечно, достаточно взять $\alpha_0 = 1 + \max_y \beta_y$.

Утверждения 1–3 придают параметру α точный смысл: он задаёт непрерывную шкалу между взвешенным усреднением из [2] и критерием max-min, защищающим наименее обеспеченный язык.

4 Численные алгоритмы

4.1 Точное решение: формулировка MILP

При $\alpha = \infty$ введением вспомогательной переменной t задача становится линейной:

$$\max\ t, \qquad U_a(x) \geqslant t, \quad a = 1, \dots, A, \qquad (7)\text{–}(10). \tag{14}$$

При $0 < \alpha < \infty$, поскольку φ_α вогнута, она аппроксимируется сверху касательными. Для вспомогательных переменных s_a и узлов $u_1 < \dots < u_L$

$$\max \sum_{a=1}^{A} w_a s_a, \qquad s_a \leqslant \varphi_\alpha(u_\ell) + \varphi'_\alpha(u_\ell)\,\big(U_a(x) - u_\ell\big), \quad \ell = 1, \dots, L, \tag{15}$$

при ограничениях (7)–(10). В эксперименте использовалось $L = 60$ равноотстоящих узлов на отрезке [0,25; 1]; найденное решение переоценивается по точной функции φ_α. Задачи решались методом ветвей и границ с помощью решателя HiGHS [12].

4.2 Приближённое решение: зеркальный подъём

Для задач большой размерности предлагается алгоритм, основанный на релаксации из утверждения 2. При $\alpha = \infty$ негладкая функция min сглаживается [13]:

$$S_\tau(u) = -\frac{1}{\tau} \ln \sum_{a=1}^{A} e^{-\tau u_a}, \qquad \min_a u_a - \frac{\ln A}{\tau} \leqslant S_\tau(u) \leqslant \min_a u_a. \tag{16}$$

На произведении симплексов шаг энтропийного зеркального подъёма [14] сводится к мультипликативному обновлению для каждого понятия:

$$x_{n,c}^{(t)} = \frac{x_{n,c}^{(t-1)} \exp\Big(\eta_t \sum_a g_a^{(t)} p_{a,n,c}\Big)}{\sum_{c' \in C_n(z)} x_{n,c'}^{(t-1)} \exp\Big(\eta_t \sum_a g_a^{(t)} p_{a,n,c'}\Big)}, \qquad \eta_t = \frac{\eta_0}{\sqrt{t}}, \tag{17}$$

где $g^{(t)}$ — нормированный градиент целевой функции по U: при $\alpha < \infty$ — $g_a \propto w_a U_a^{-\alpha}$, при $\alpha = \infty$ — $g_a \propto e^{-\tau_t\left(U_a - \min\limits_b U_b\right)}$.

Алгоритм 1 Приближённое решение задачи (P_α)

1: Для каждой комбинации правил $z \in R_1 \times \dots \times R_K$ определить допустимые множества $C_n(z)$ и положить $x_{n,c}^{(0)} = 1/|C_n(z)|$.
2: Для $t = 1, \dots, T$: вычислить $u^{(t)} = U(x^{(t-1)})$, найти $g^{(t)}$ и выполнить обновление по (17); при $\alpha = \infty$ температура τ_t увеличивается в геометрической прогрессии от τ_0 до τ_1.
3: Округление: $\widehat{c}_n = \arg\max\limits_c x_{n,c}^{(T)}$.
4: Локальный поиск: выполнять замены в пределах одного понятия, увеличивающие целевую функцию (при $\alpha = \infty$ — лексикографически увеличивающие упорядоченный вектор U), пока такие замены существуют.
5: Выбрать наилучшее решение среди всех z.

Одна итерация требует $O\bigl(A\sum_n |C_n|\bigr)$ операций. Для вогнутой липшицевой функции погрешность энтропийного зеркального подъёма за T итераций убывает как $O\bigl(G\sqrt{\sum_n \ln|C_n|}\,/\sqrt{T}\bigr)$ [14], где G — оценка нормы градиента; при $\alpha = \infty$ к ней добавляется погрешность сглаживания $\ln A/\tau$ из (16).

5 Вычислительный эксперимент

5.1 Данные и параметры

В эксперименте использовалась CLDF-версия лексической базы NorthEuraLex 0.9 [11]. Из базы были взяты турецкий, азербайджанский, узбекский, казахский и татарский языки; все 1016 понятий представлены во всех пяти языках с транскрипцией IPA. Число кандидатов составляет 5770 (в среднем 5,68 на понятие, не более 12). Веса языков взяты из [2]: $w = (0{,}346;\ 0{,}195;\ 0{,}180;\ 0{,}146;\ 0{,}133)$. Ввиду отсутствия данных о частотах принято $f_n = 1/N$.

Рассматривались два класса соответствий: начальные заднеязычные согласные (g, ɟ или k, c) и зубные согласные (d или t). Если два таких кандидата для одного понятия были схожи по части, следующей за начальным согласным ($p \geqslant 0{,}6$), они помечались как зависящие от соответствующего правила. В результате от правил оказались зависимыми 474 кандидата в 168 понятиях.

Вычисления выполнены с использованием Python 3.13, NumPy 2.5 и SciPy 1.18 [15] (решатель HiGHS). Параметры алгоритма 1: $\eta_0 = 2$; при $\alpha < \infty$ — $T = 1500$; при $\alpha = \infty$ — $T = 3000$, $\tau_0 = 20$, $\tau_1 = 400$.

5.2 Влияние параметра α

В таблице 1 приведены точные решения, полученные с ограничениями согласованности.

Таблица 1 Понятность U_a по языкам при различных значениях α ($\bar{U}$ — невзвешенное среднее).

α	турецкий	азербайджанский	узбекский	казахский	татарский	$\sum_a w_a U_a$	$\bar{U}$	$\min\limits_a U_a$
0	0,936	0,795	0,645	0,528	0,574	0,748	0,696	0,528
1	0,876	0,789	0,689	0,564	0,612	0,745	0,706	0,564
2	0,841	0,777	0,702	0,593	0,630	0,739	0,709	0,593
4	0,790	0,752	0,713	0,633	0,659	0,728	0,709	0,633
∞	0,704	0,704	0,704	0,704	0,704	0,704	0,704	0,704

При взвешенном усреднении ($\alpha = 0$) норма сильно смещается в сторону турецкого языка: в 822 из 1016 понятий (81 %) выбирается турецкая форма, а понятность для наименее обеспеченного языка — казахского — равна 0,528. С ростом α понятность для наименее обеспеченного языка монотонно возрастает и в max-min решении становится практически одинаковой для всех пяти языков — 0,704. При этом источники выбранных форм уравновешиваются: казахский — 296, турецкий — 235, узбекский — 212, татарский — 201, азербайджанский — 140 понятий (одна форма может встречаться в нескольких языках). Решения при $\alpha = 0$ и max-min различаются в 675 понятиях.

На рис. 1 приведена граница Парето, построенная методом ε-ограничений: каждая точка является максимумом взвешенного среднего при условии $\min\limits_a U_a \geqslant \varepsilon$. «Цену справедливости» [3] можно оценить следующим образом: повышение понятности для наименее обеспеченного языка с 0,528 до 0,600 снижает взвешенное среднее лишь

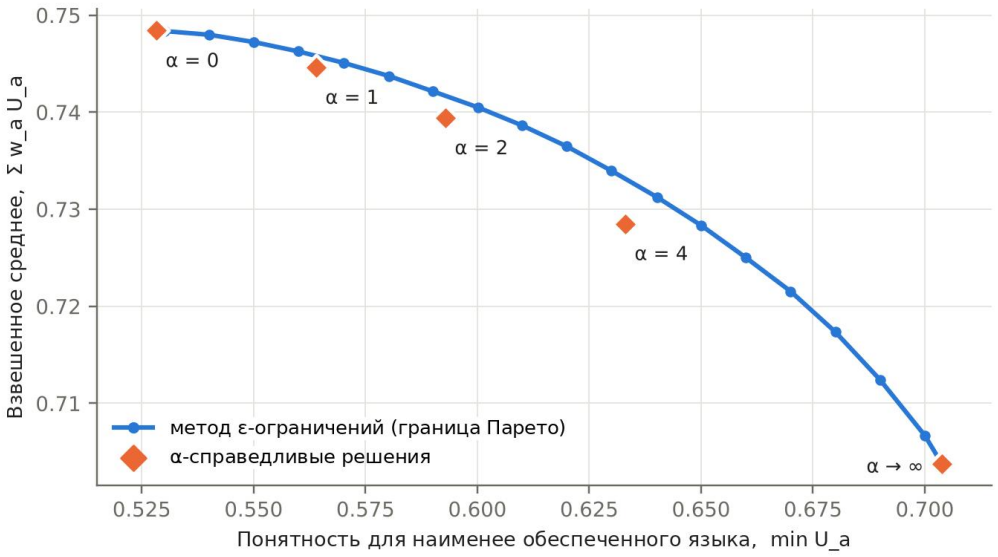


Рис. 1 Граница Парето: компромисс между понятностью для наименее обеспеченного языка и взвешенным средним.

на 1,1 %, а до 0,704 — на 6,0 %, тогда как выигрыш наименее обеспеченного языка составляет 33,2 %. Невзвешенная средняя понятность $\bar{U}$ даже возрастает на 1,2 %. α-справедливые решения располагаются несколько ниже кривой, поскольку учитывают распределение по всем языкам, а не только наименее обеспеченный язык.

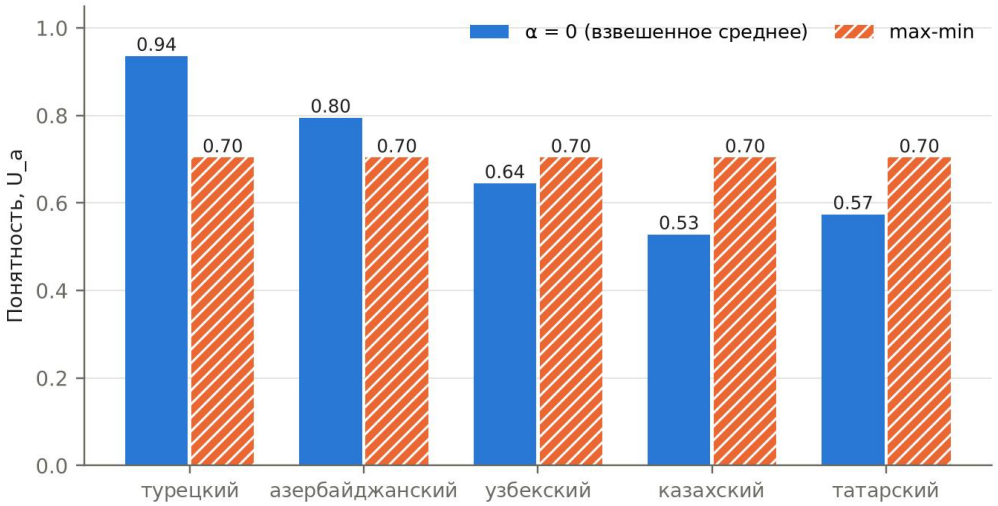


Рис. 2 Понятность по языкам в решениях при $\alpha = 0$ и max-min.

Даже при равных весах языков ($w_a = 0{,}2$) в решении при $\alpha = 0$ наименее обеспеченным остаётся казахский язык ($\min\limits_a U_a = 0{,}602$, рис. 2). Следовательно, дисбаланс обусловлен не только весами, но и самим критерием усреднения.

5.3 Влияние ограничений согласованности

В решениях без ограничений в обоих классах соответствий оба правила применяются вперемешку: например, при $\alpha = 0$ в начале слова k выбирается в 68 понятиях, а g — в 21. Ограничения (8)–(9) полностью устраняют это противоречие; при этом

при $\alpha = 0$ взвешенное среднее снижается с 0,7503 до 0,7484 (на 0,25 %), а в max-min решении понятность для наименее обеспеченного языка — с 0,7040 до 0,7038 (на 0,03 %). Следовательно, согласованная норма достигается практически без потерь. Справедливые решения ($\alpha \geqslant 1$) в обоих классах выбирают правила с глухими согласными (k, t).

5.4 Сравнение алгоритмов

В таблице 2 сопоставлены результаты MILP и алгоритма 1.

Таблица 2 Сравнение точного (MILP) и приближённого (алгоритм 1) решений.

α	MILP: F_α	MILP: время, с	Алгоритм 1: F_α	Алгоритм 1: время, с
1	−0,30811	175,1	−0,30810	3,6
2	−1,37572	75,5	−1,37572	3,8
4	−0,89899	72,5	−0,89893	4,2
∞	0,70375	8,6	0,70356	7,0

При $0 < \alpha < \infty$ алгоритм 1 даёт практически те же результаты, что и MILP, и даже несколько лучшие. Это объясняется аппроксимацией касательными в (15), тогда как алгоритм 1 работает с точной функцией φ_α. Время вычислений меньше в 17–49 раз. При $\alpha = \infty$ погрешность алгоритма 1 составляет 0,03 %. Точность решения MILP подтверждается оценкой (12): значение LP-релаксации равно 0,70379, т. е. отклонение от оптимума не превышает 0,006 %.

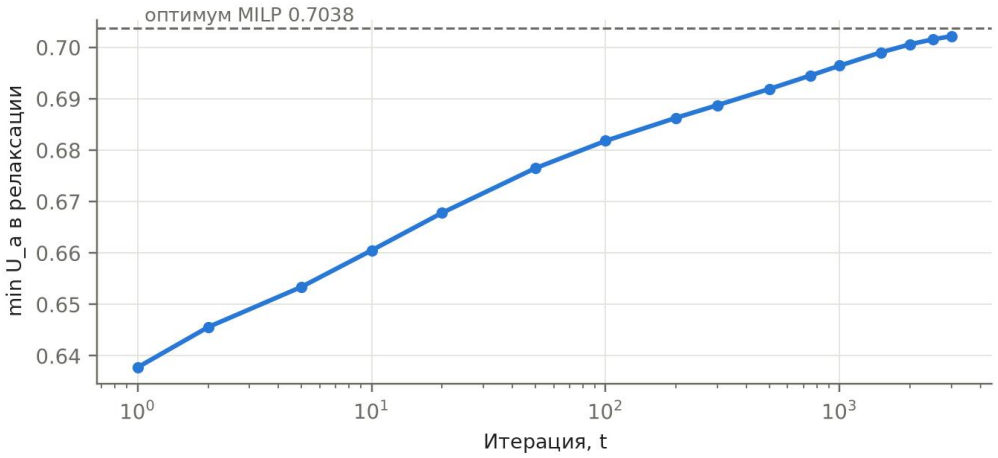


Рис. 3 Сходимость алгоритма 1 ($\alpha = \infty$; понятность для наименее обеспеченного языка в релаксации).

Из рис. 3 видно, что $\min\limits_a U_a$ в релаксации растёт практически линейно относительно логарифма числа итераций, и после 3000 итераций отклонение от оптимума составляет менее 0,25 %; оставшееся отклонение устраняется округлением и локальным поиском.

5.5 Примеры

В таблице 3 приведены некоторые понятия, для которых решения при $\alpha = 0$ и max-min различаются.

Таблица 3 Примеры выбранных форм (IPA; в скобках указан язык-источник).

Понятие	$\alpha = 0$	max-min
море	dɛniz (тур.)	teŋəz (каз.)
горький	adʒɯ (тур.)	atɕɤ (тат.)
нести	dɑʃɯmɑg (азерб.)	taʃɯmak (тур.)
тридцать	otuz (тур.)	otəz (каз.)
они	onlaɾ (тур.)	olɑɾ (каз.)
коготь	dɯrnɑg (азерб.)	tɤrnak (тат.)

Примеры наглядно демонстрируют действие ограничения согласованности: в справедливом решении в начале слова вместо d выбирается t (dɛniz $\to$ teŋəz, dɑʃɯmɑg $\to$ taʃɯmak).

6 Обсуждение и ограничения

Полученные результаты следует интерпретировать с учётом следующих ограничений. Во-первых, показатель (1) является косвенной оценкой (прокси) вероятности понимания и ещё не откалиброван по экспериментальным данным, полученным от носителей языков; на следующем этапе планируется оценить логистическую модель вида $\pi_{a,n,c} = \sigma\left(\theta_0 + \theta_1 p_{a,n,c} + \cdots\right)$ на основе тестов на перевод слов. Во-вторых, формы NorthEuraLex являются словарными; в частности, глаголы приведены с аффиксами инфинитива (турецкий *-mak*, казахский *-u*), что влияет на сходство глаголов; целесообразен анализ на уровне основ. В-третьих, частоты приняты равными и рассмотрены только два класса соответствий. Указанные ограничения не затрагивают математическую структуру модели: откалиброванные $p_{a,n,c}$, корпусные частоты и дополнительные правила включаются в задачу (6)–(10) без изменений.

7 Заключение

Задача выбора общей лексической нормы тюркских языков поставлена как задача α-справедливой дискретной оптимизации с ограничениями фонологической согласованности. Доказано, что модель взвешенного усреднения является её частным случаем при $\alpha = 0$, а критерий max-min — её пределом при $\alpha \to \infty$. Разработаны формулировка задачи в виде MILP, оценка точности на основе релаксации и приближённый алгоритм на основе метода зеркального подъёма. Эксперимент на 1016 понятиях пяти тюркских языков показал, что взвешенное усреднение смещает норму в сторону турецкого языка (81 % форм), тогда как справедливое решение повышает понятность для наименее обеспеченного языка на 33,2 %, снижая взвешенное среднее лишь на 6,0 %. Фонологическая согласованность обеспечивается практически без потерь (0,03–0,25 %). Предложенный алгоритм отличается от точного решения менее чем на 0,03 % и работает до 49 раз быстрее.

Дальнейшие исследования будут направлены на экспериментальную калибровку вероятности понимания, расширение числа языков (башкирский, киргизский, туркменский, уйгурский) и моделирование динамики распространения нормы по регионам [16].

Литература

[1] *Karimov B. R., Mutalov Sh. Sh.* O‘rtaturk tili. — Toshkent: Mehnat, 1992.

[2] *Boborahimov B. I., Axmedjanova D. A., Sharipov Kh. D.* A corpus-based model of the Middle Turkic language using weighted averaging // Problems of Computational and Applied

Mathematics. — 2025. — № 5(69). — P. 123–131. — doi: https://doi.org/10.71310/pcam.5_69.2025.10.

[3] *Bertsimas D., Farias V. F., Trichakis N.* The price of fairness // Operations Research. — 2011. — Vol. 59, № 1. — P. 17–31.

[4] *Rawls J.* A Theory of Justice. — Cambridge, MA: Harvard University Press, 1971.

[5] *Mo J., Walrand J.* Fair end-to-end window-based congestion control // IEEE/ACM Transactions on Networking. — 2000. — Vol. 8, № 5. — P. 556–567.

[6] *Nash J. F.* The bargaining problem // Econometrica. — 1950. — Vol. 18, № 2. — P. 155–162.

[7] *Gooskens C.* The contribution of linguistic factors to the intelligibility of closely related languages // Journal of Multilingual and Multicultural Development. — 2007. — Vol. 28, № 6. — P. 445–467.

[8] *Heeringa W.* Measuring dialect pronunciation differences using Levenshtein distance: PhD thesis. — Groningen: University of Groningen, 2004.

[9] *Needleman S. B., Wunsch C. D.* A general method applicable to the search for similarities in the amino acid sequence of two proteins // Journal of Molecular Biology. — 1970. — Vol. 48, № 3. — P. 443–453.

[10] *Kondrak G.* A new algorithm for the alignment of phonetic sequences // Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL). — Seattle, 2000. — P. 288–295.

[11] *Dellert J., Daneyko T., Münch A. [et al.]* NorthEuraLex: a wide-coverage lexical database of Northern Eurasia // Language Resources and Evaluation. — 2020. — Vol. 54. — P. 273–301. — doi: https://doi.org/10.1007/s10579-019-09480-6.

[12] *Huangfu Q., Hall J. A. J.* Parallelizing the dual revised simplex method // Mathematical Programming Computation. — 2018. — Vol. 10, № 1. — P. 119–142.

[13] *Nesterov Yu.* Smooth minimization of non-smooth functions // Mathematical Programming. — 2005. — Vol. 103, № 1. — P. 127–152.

[14] *Beck A., Teboulle M.* Mirror descent and nonlinear projected subgradient methods for convex optimization // Operations Research Letters. — 2003. — Vol. 31, № 3. — P. 167–175.

[15] *Virtanen P., Gommers R., Oliphant T. E. [et al.]* SciPy 1.0: fundamental algorithms for scientific computing in Python // Nature Methods. — 2020. — Vol. 17. — P. 261–272.

[16] *Abrams D. M., Strogatz S. H.* Modelling the dynamics of language death // Nature. — 2003. — Vol. 424. — P. 900.

UDC 519.85:81'322

AN α-FAIR OPTIMIZATION MODEL AND NUMERICAL ALGORITHMS FOR SELECTING A COMMON LEXICAL NORM OF TURKIC LANGUAGES

[1*]***Bakayev I.I.,*** [2]***Akhmedjanova D.A.***

[*]i.i.bakaev@buxdu.uz, [2]daxmadjonova@gmail.com

[1]Bukhara State University, 11, M. Iqbol Street, Bukhara, 200114 Uzbekistan;

[2]Research Institute for the Development of Digital Technologies and Artificial Intelligence, 17A, Buz-2, Tashkent, 100125 Uzbekistan

The selection of a common lexical norm for Turkic languages is formulated as a discrete optimization problem that ensures fair intelligibility for speakers of all languages.

For each concept, one form is chosen from a set of candidates; intelligibility for speakers of each language is estimated by a feature-weighted edit distance, and the choice is governed by an α-fair objective function together with phonological consistency constraints. We prove that the weighted averaging model is the special case $\alpha = 0$, while the max-min criterion is the limit $\alpha \to \infty$. An exact mixed-integer linear programming formulation and an approximate algorithm based on relaxation, mirror ascent, rounding and local search are developed. A computational experiment on 1016 concepts of five Turkic languages from the NorthEuraLex database shows that under weighted averaging the intelligibility for the worst-off language (Kazakh) is 0.528, whereas the max-min solution raises it to 0.704 (by 33.2%) while reducing the population-weighted average by only 6.0%. The proposed algorithm deviates from the exact solution by less than 0.03% and is up to 49 times faster for $\alpha = 1$.

Keywords: Turkic languages, common norm, α-fairness, max-min optimization, mixed-integer programming, mirror ascent, phonetic distance, NorthEuraLex

Citation: Bakayev I.I., Akhmedjanova D.A. 2026. An α-fair optimization model and numerical algorithms for selecting a common lexical norm of Turkic languages. *Problems of Computational and Applied Mathematics.* 5(76): 137-147.

DOI: 10.71310/pcam.5_76.2026.09

№ 5(76) 2026

ISSN 2181-8460

HISOBLASH VA AMALIY MATEMATIKA MUAMMOLARI

ПРОБЛЕМЫ ВЫЧИСЛИТЕЛЬНОЙ И ПРИКЛАДНОЙ МАТЕМАТИКИ

PROBLEMS OF COMPUTATIONAL AND APPLIED MATHEMATICS

AiRI
RAQAMLI TEXNOLOGIYALAR VA SUN'IY INTELLEKTNI RIVOJLANTIRISH ILMIY-TADQIQOT INSTITUTI

ПРОБЛЕМЫ ВЫЧИСЛИТЕЛЬНОЙ И ПРИКЛАДНОЙ МАТЕМАТИКИ

№ 5(76) 2026

Журнал основан в 2015 году.
Издается 6 раз в год.

Учредитель:
Научно-исследовательский институт развития цифровых технологий и искусственного интеллекта.

Главный редактор:
Равшанов Н.

Заместители главного редактора:
Арипов М.М., Шадиметов Х.М., Ахмедов Д.Д.

Ответственный секретарь:
Убайдуллаев М.Ш.

Редакционный совет:
Азамов А.А., Алоев Р.Д., Амиргалиев Е.Н. (Казахстан), Арушанов М.Л., Абдуллозода С.Ф. (Таджикистан), Боборахимов Б.И., Бурнашев В.Ф., Джумаёзов У.З., Загребина С.А. (Россия), Задорин А.И. (Россия), Игнатьев Н.А., Ильин В.П. (Россия), Иманкулов Т.С. (Казахстан), Исмагилов И.И. (Россия), Кабанихин С.И. (Россия), Курбонов Н.М., Маматов Н.С., Мирзаев Н.М., Мурадов Ф.А., Назирова Э.Ш., Нормуродов Ч.Б., Нуралиев Ф.А., Нуралиев Ф.М., Опанасенко В.Н. (Украина), Расулмухамедов М.М., Садуллаева Ш.А., Старовойтов В.В. (Беларусь), Хаётов А.Р., Халджигитов А., Хамдамов Р.Х., Хужаев И.К., Хужаеров Б.Х., Эшмаматова Д.Б., Дустмуродова Ш.Ж., Чье Ен Ун (Россия), Шабозов М.Ш. (Таджикистан), Dimov I. (Болгария), Li Y. (США), Mascagni M. (США), Min A. (Германия), Singh M. (Южная Корея).

Журнал зарегистрирован в Агентстве информации и массовых коммуникаций при Администрации Президента Республики Узбекистан.
Свидетельство №0856 от 5 августа 2015 года.

ISSN 2181-8460, eISSN 2181-046X

При перепечатке материалов ссылка на журнал обязательна.
За точность фактов и достоверность информации ответственность несут авторы.

Адрес редакции:
100125, г. Ташкент, м-в. Буз-2, 17А.
Тел.: +(998) 71 263-41-98.
Э-почта: journals@airi.uz.
Веб-сайт: https://journals.airi.uz.

Дизайн и вёрстка:
Шарипов Х.Д.

Отпечатано в типографии НИИ РЦТИИ.
Подписано в печать 28.10.2026 г.
Формат 60x84 1/8. Заказ № 6. Тираж 100 экз.

PROBLEMS OF COMPUTATIONAL AND APPLIED MATHEMATICS

No. 5(76) 2026

The journal was established in 2015.
6 issues are published per year.

Founder:
Digital Technologies and Artificial Intelligence Development Research Institute.

Editor-in-Chief:
Ravshanov N.

Deputy Editors:
Aripov M.M., Shadimetov Kh.M., Akhmedov D.D.

Executive Secretary:
Ubaydullaev M.Sh.

Editorial Council:
Azamov A.A., Aloev R.D., Amirgaliev E.N. (Kazakhstan), Arushanov M.L., Abdullozoda S.F. (Tajikistan), Boborakhimov B.I., Burnashev V.F., Djumayozov U.Z., Zagrebina S.A. (Russia), Zadorin A.I. (Russia), Ignatiev N.A., Ilyin V.P. (Russia), Imankulov T.S. (Kazakhstan), Ismagilov I.I. (Russia), Kabanikhin S.I. (Russia), Kurbonov N.M., Mamatov N.S., Mirzaev N.M., Muradov F.A., Nazirova E.Sh., Normurodov Ch.B., Nuraliev F.A., Nuraliev F.M., Opanasenko V.N. (Ukraine), Sadullaeva Sh.A., Starovoitov V.V. (Belarus), Khayotov A.R., Khaldjigitov A., Khamdamov R.Kh., Khujaev I.K., Khujayorov B.Kh., Eshmamatova D.B., Dustmurodova Sh.J., Chye En Un (Russia), Shabozov M.Sh. (Tajikistan), Dimov I. (Bulgaria), Li Y. (USA), Mascagni M. (USA), Min A. (Germany), Singh M. (South Korea).

The journal is registered by Agency of Information and Mass Communications under the Administration of the President of the Republic of Uzbekistan.
Certificate of Registration No. 0856 of 5 August 2015.

ISSN 2181-8460, eISSN 2181-046X

At a reprint of materials the reference to the journal is obligatory.
Authors are responsible for the accuracy of the facts and reliability of the information.

Address:
100125, Tashkent, Buz-2, 17A.
Tel.: +(998) 71 263-41-98.
E-mail: journals@airi.uz.
Web-site: https://journals.airi.uz.

Layout design:
Sharipov Kh.D.

DTAIRI printing office.
Signed for print 28.10.2026
Format 60x84 1/8. Order No. 6. Print run of 100 copies.

Содержание

Расулов А.С., Раимова Г.М., Рахматов М.Й.
Метод Монте-Карло для решения нелинейного кинетического уравнения Больцмана 8

Эшкулов М.У., Хамдамов Р.Х.
Многокритериальная оптимизация системы подъёма воды с использованием солнечных фотоэлектрических панелей 22

Абдуллаев К.Д.
Принцип разделения классов и зависимость результатов кластеризации от меры расстояния 33

Рамазонов Ш.Ш.
Селекция объектов по отношениям с ближайшими объектами противоположного класса и сохранение локальной структуры обучающей выборки . . 46

Исомиддинов А.И., Акбарова А.А., Акбаров Б.Х.
Марковская переключающаяся векторная авторегрессия для скрытых гликемических режимов в реальных данных самоконтроля диабета: оценка на отложенной выборке 60

Ким К.К.
Сравнительный анализ XAI-подходов для обнаружения синтетической речи 72

Эшмаматова Д.Б., Солижанова К.А.
Пфаффианова структура и спектральная динамика дискретного оператора Лотки–Вольтерры: подход к интеграции геологических данных 95

Джумаёзов У.З., Абдирахмонова М.Н., Усмонов Л.С.
Численное решение трёхмерной связанной динамической задачи термоупругости в деформациях 119

Бакаев И.И., Ахмеджанова Д.А.
α-справедливая оптимизационная модель и численные алгоритмы выбора общей лексической нормы тюркских языков 137

Равшанов Н., Камолиддинова Ф., Lu Y.
Моделирование переноса и диффузии загрязняющих веществ в пограничном слое атмосферы с учётом орографии местности и метеорологических факторов 148

Contents

Rasulov A.S., Raimova G.M., Rakhmatov M.Y.
Monte Carlo solution of the nonlinear kinetic Boltzmann equation 8
Eshkulov M.U., Khamdamov R.Kh.
Multi-criteria optimization of solar photovoltaic-based water lifting systems . . . 22
Abdullaev K.D.
The principle of class separation and the dependence of clustering results on the distance measure 33
Ramazonov Sh.Sh.
Instance selection based on nearest-enemy relations and preservation of the local structure of a training sample 46
Isomiddinov A.I., Akbarova A.A., Akbarov B.X.
Markov-switching vector autoregression for latent glycaemic regimes in real-world diabetes self-monitoring data: an out-of-sample evaluation 60
Kim K.K.
Comparative analysis of XAI approaches for synthetic speech detection 72
Eshmamatova D.B., Solijanova K.A.
Pfaffian structure and spectral dynamics of a discrete Lotka–Volterra operator: a framework for geological evidence integration 95
Djumayozov U.Z., Abdirakhmonova M.N., Usmonov L.S.
Numerical solution of a three-dimensional coupled dynamic thermoelasticity problem in terms of strains 119
Bakayev I.I., Akhmedjanova D.A.
An α-fair optimization model and numerical algorithms for selecting a common lexical norm of Turkic languages 137
Ravshanov N., Kamoliddinova F., Lu Y.
Modeling pollutant transport and diffusion in the atmospheric boundary layer considering terrain orography and meteorological factors 148