

Задача 2

Даны предложения на русском языке и их переводы на язык дирбал* в перепутанном порядке:

- | | |
|--------------------------------|--|
| 1. Вы пришли и побили нас. | A. nana baniju purra balgan |
| 2. Вы побили нас, и мы пришли. | B. nana purrana balgan nana baniju |
| 3. Мы побили вас и пришли. | C. purra baniju purra nanana balgan |
| 4. Мы пришли, и вы побили нас. | D. purra nanana balgan baniju |

Задание 1. Установите правильные соответствия.

Нетрудно заметить, что **baniju** и **balgan** – это глаголы, а **nana**, **nanana**, **purra** и **purrana** – местоимения. Однако, сосчитав количество местоимений в русских предложениях и их переводов на язык дирбал, находим несовпадения.

Мы – 3, нас – 3 (вместе 6).	nana – 3, nanana – 2 (вместе 5).
Вы – 3, вас – 1 (вместе 4).	purra – 4, purrana – 1 (вместе 5).

Это может навести на мысль, что в языке дирбал местоимения используются не так, как в русском.

В русском языке возможно опущение одинаковых местоимений, превращающее сложносочинённое предложение в простое предложение с однородными членами.

Например, вместо «вы пришли, и вы побили нас» говорим «вы пришли и побили нас».

Восстановим опущенные местоимения в русских предложениях. Получаем:

Вы пришли, и **ВЫ** побили нас.
Вы побили нас, и мы пришли.
Мы побили вас, и **МЫ** пришли.
Мы пришли, и вы побили нас.

Заново считаем количество, получаем:

Мы – 4, нас – 3 (вместе 7).	nana – 3, nanana – 2 (вместе 5).
Вы – 4, вас – 1 (вместе 5).	purra – 4, purrana – 1 (вместе 5).

Теперь получается, что местоимений в русских предложениях – 12, а в предложениях на языке дирбал – 10.

Следовательно, можно предположить, что в языке дирбал тоже могут опускаться какие-либо из местоимений, только не те, что в русском.

Взглянув на количество разных местоимений в обоих языках, можем заметить, что «вы» и «вас» встречаются ровно столько же раз, сколько и **purra** и **purrana**. Тогда можно предположить, что местоимение «вы» – **purra**, а «вас» – **purrana**. В таком случае, «мы» – **nana**, а «нас» – **nanana**. Получается, что по одному разу опускаются местоимения «мы» и «нас», но с этим разберёмся позже, т.к. для выполнения Задания 1 информации теперь достаточно.

Задание 1. Установите правильные соответствия:

Вы пришли и побили нас.	=	purra banıju purra nanana balgan (дословно: <i>вы пришли вы нас побили</i>)
Вы побили нас, и мы пришли.	=	purra nanana balgan banıju (дословно <i>вы нас побили пришли</i>)
Мы побили вас и пришли.	=	nana purrana balgan nana banıju (дословно <i>мы вас побили мы пришли</i>)
Мы пришли, и вы побили нас.	=	nana banıju purra balgan (дословно <i>мы пришли вы побили</i>)

Задание 2. Переведите на русский:

nana purrana balgan banıju дословно может переводиться как *мы вас побили пришли*.

Пока ещё не понятно, кто пришёл: мы пришли или вы пришли. На то, что правильный ответ именно «мы вас побили, и вы пришли», намекает тот факт, что предложение «мы вас побили и пришли» нам уже дано, и оно на языке дирбал выглядит по-другому.

Это предположение можно проверить, взглянув на то, какие местоимения опускаются.

В переводе предложения «вы побили нас, и мы пришли» пропущено «мы», а в переводе предложения «мы пришли, и вы побили нас» пропущено «нас».

Мы (nana) пропущено в предложении, где уже упоминалось слово «нас» (nanana), а «нас» пропущено в предложении, где уже упоминалось слово «мы».

Тогда получается, что **если прямое дополнение одной половины предложения выражено тем же местоимением, что и подлежащее другого, то опускается второе местоимение из этой пары.**

В таком случае и предположение, что в предложении nana purrana balgan banıju (дословно – «мы вас побили пришли») пропущено именно местоимение «вы», подтверждается, т.к. в первой части уже есть местоимение «вас».

nana purrana balgan banıju = Мы побили вас, и вы пришли.

В задаче показано явление т.н. синтаксической эргативности. В языке дирбал можно пропускать местоимения таким образом, потому что, в отличие от русского, где подлежащее предложения с непереходным глаголом (например, «мы пришли») группируется (и всячески ведёт себя похожим образом) с подлежащим предложения с переходным глаголом (например, «мы ударили вас»), в языках, в которых проявляется синтаксическая эргативность, подлежащее предложения с непереходным глаголом группируется не с подлежащим, а с дополнением предложения с переходным глаголом.

Другими словами, в русском языке предложение, выглядящее как «Я увидел дядю Стёпу и поздоровался» означает «Я увидел дядю Стёпу, и я поздоровался», а в языке дирбал – «Я увидел дядю Стёпу, и он поздоровался».

Теперь без труда можем выполнить задание 3.

Вы пришли, и мы побили вас.	=	purra banıju nana balgan (дословно <i>вы пришли мы побили</i>)
Вы побили нас и пришли.	=	purra nanana balgan purra banıju (дословно <i>вы нас побили вы пришли</i>)

Задача 3

Даны даты на языке сольресоль** и их переводы в перепутанном порядке:

- | | |
|------------------------------------|------------------------|
| 1. solsolfa, rere mi mimifa | A. суббота, 20 января |
| 2. solsolfa, reresi mifafa | B. среда, 3 мая |
| 3. solsolmi, dolala refafa | C. среда, 8 августа |
| 4. solsolmi, rere mi mimire | D. четверг, 9 августа |
| 5. solsolsi, dorere fafare | E. суббота, 17 августа |
| 6. solsolsi, rere mi milala | F. четверг, 15 декабря |

Задание 1. Установите правильные соответствия.

Три раза встречается слово 'rere mi', а три раза слово «август», из чего делаем соответствующий вывод.

Имеем 8, 9 и 17 августа, которым соответствуют **rere mi mimire**, **rere mi mimifa** и **rere mi milala**. Если предположить, что слова, отличающиеся одной нотой, **mimire** и **mimifa**, – это 8 и 9, то клубок распутывается быстро:

- | | |
|------------------------------------|------------------------|
| 1. solsolfa, rere mi mimifa | D. четверг, 9 августа |
| 2. solsolfa, reresi mifafa | F. четверг, 15 декабря |
| 3. solsolmi, dolala refafa | B. среда, 3 мая |
| 4. solsolmi, rere mi mimire | C. среда, 8 августа |
| 5. solsolsi, dorere fafare | A. суббота, 20 января |
| 6. solsolsi, rere mi milala | E. суббота, 17 августа |

Для того, чтобы решить оставшиеся задания, составим таблицу из имеющихся месяцев и дней недели.

Jan	dorere		Mon	
Feb			Tue	
Mar			Wed	solsolmi
Apr			Thu	solsolfa
May	dolala		Fri	
Jun			Sat	solsolsi
Jul			Sun	
Aug	rere mi			
Sep				
Oct				
Nov				
Dec	reresi			

Видно, что названия месяцев и дней недели состоят из двух частей: «нота1-нота2нота2» или «нота1нота1-нота2».

Причём, чем выше нота1 и нота2 в гамме do-re-mi-fa-sol-la-si, тем более поздний месяц. При этом нота2 увеличивается с шагом 1.

Заполним пропуски.

Jan	dorere		Mon	<i>solsoldo</i>
Feb	<i>domimi</i>		Tue	<i>solsolre</i>
Mar	<i>dofafa</i>		Wed	solsolmi
Apr	<i>dosolsol</i>		Thu	solsolfa
May	dolala		Fri	
Jun	<i>dosisi</i>		Sat	solsolsi
Jul			Sun	
Aug	reremi			
Sep	<i>rerefa</i>			
Oct	<i>reresol</i>			
Nov	<i>rerela</i>			
Dec	reresi			

На место июля годится как **reredo**, так и **rerere**, а на место пятницы – как **solsolsol**, так и **solsolla**.

Логично предположить, что комбинация из трёх идентичных нот (другими словами, случай, когда нота1 = нота2) пропускается. Это гипотезу подтверждает тот факт, что месяцы начинаются с **dorere**, а не с ***dododo**.

Jan	dorere		Mon	<i>solsoldo</i>
Feb	<i>domimi</i>		Tue	<i>solsolre</i>
Mar	<i>dofafa</i>		Wed	solsolmi
Apr	<i>dosolsol</i>		Thu	solsolfa
May	dolala		Fri	<i>solsolla</i>
Jun	<i>dosisi</i>		Sat	solsolsi
Jul	<i>reredo</i>		Sun	?
Aug	reremi			
Sep	<i>rerefa</i>			
Oct	<i>reresol</i>			
Nov	<i>rerela</i>			
Dec	reresi			

Пока непонятно, что делать с воскресеньем, но с этим разберёмся позже. Составим таблицу из имеющихся числительных.

1			11	
2			12	
3	re-fafa		13	
4			14	
5			15	mi-fafa
6			16	
7			17	mi-lala
8	mimi-re		18	
9	mimi-fa		19	
10			20	fafa-re

Логично предположить, что образование числительных происходит похожим образом, что и образование месяцев и дней недели.

1	<i>re-dodo</i>		11	<i>mimi-la</i>
2	<i>re-mimi</i>		12	<i>mimi-si</i>
3	re-fafa		13	<i>mi-dodo</i>
4	<i>re-solsol</i>		14	<i>mi-rere</i>
5	<i>re-lala</i>		15	mi-fafa
6	<i>re-sisi</i>		16	<i>mi-solsol</i>
7	<i>mimi-do</i>		17	mi-lala
8	mimi-re		18	<i>mi-sisi</i>
9	mimi-fa		19	<i>fafa-do</i>
10	<i>mimi-sol</i>		20	fafa-re

Отсюда яснее видно, что числительные образуются циклами по 12 с подциклами по 6: нота1 сначала шесть раз предстаёт в удвоенном, а потом шесть раз – в неудвоенном виде, после чего то же самое повторяется со следующей нотой1.

Теперь можем вернуться к воскресению.

Mon	<i>solsoldo</i>
Tue	<i>solsolre</i>
Wed	solsolmi
Thu	solsolfa
Fri	<i>solsolla</i>
Sat	solsolsi
Sun	<i>soldodo</i>

Т.к. нота1 сначала предстаёт в удвоенном виде, а потом в неудвоенном, а не наоборот, воскресенье – это **soldodo** (а не ***ladodo**, как можно было бы ошибочно предположить).

Теперь у нас есть достаточно данных, чтобы выполнить оставшиеся задания.

Задание 2.

- | | |
|--------------------------------------|---------------------|
| 7. solsolre, reresol resolsol | вторник, 4 октября |
| 8. solsolla, domimi midodo | пятница, 13 февраля |

Задание 3.

- | | |
|---------------------------|--------------------------------|
| 9. понедельник, 6 июля | solsoldo, reredo resisi |
| 10. воскресенье, 19 марта | soldodo, dofafa fafado |
-

Задача 4

В автоматической обработке естественного языка (например, при автоматической проверке орфографии), часто бывает нужно определить, насколько различны два написанных слова. Одна из количественных мер, используемых для этого, называется расстоянием Дамерау–Левенштейна.

Расстояние названо в честь Владимира Левенштейна и Фредерика Дамерау. Левенштейн придумал способ измерения «расстояний» между словами, а Дамерау независимо от него выделил несколько классов, в которые попадает большинство опечаток.

Даны пары английских слов и расстояние Дамерау-Левенштейна между словами каждой пары. Некоторые числа пропущены.

Пара слов		Расстояние Дамерау-Левенштейна
1. acre	car	2
2. anteater	theatre	4
3. banana	nanny	3
4. cat	crate	2
5. cocoon	cuckoo	3
6. emporium	empower	4
7. goer	ogre	2
8. lyra	lay	2
9. life	death	5
10. point	sirloin	5
11. stone	sonnet	3
12. surge	ruse	3
13. task	tusk	1
14. ape	ea	
15. baba	arab	
16. contest	toner	
17. eel	lee	
18. martial	marital	
19. monarchy	democracy	
20. seatback	backseat	
21. warfare	farewell	
22. peat	tape	
23. smoking	hospital	

Задание 1. Заполните пропуски.

Задание 2. Дайте определение расстоянию Дамерау-Левенштейна и предположите, какие классы опечаток выделил Дамерау

Задание 3. Даны два слова с длинами m и n ($m > n$). Каково максимально возможное расстояние Дамерау-Левенштейна между этими словами? Минимально возможное? (Выразите ответы через m и n).

Начнем с задания 2. Расстояние Дамерау–Левенштейна между двумя словами — это минимальное количество элементарных операций, которые необходимо совершить, чтобы получить одно слово из другого. Элементарные операции таковы: удаление одного символа; добавление одного символа; замена одного символа другим; перестановка двух соседних символов.

Соответственно, четыре класса опечаток, которые выделил Дамерау, разрабатывая алгоритм их автоматического исправления, — это пропуск символа; лишний символ; замена символа на другой; перепутанный порядок двух соседних символов.

Иногда решатели, выполняя задание 2, пытаются вместо определения изложить алгоритм, который позволял бы находить расстояние Дамерау–Левенштейна. Это, однако, довольно сложно, и для решения данной задачи совершенно не необходимо, достаточно дать определение в духе изложенного выше.

Теперь мы можем выполнить задание 1:

	Пара слов		Расстояние Дамерау–Левенштейна
15.	baba	arab	3
16.	contest	toner	4
17.	eel	lee	2
18.	martial	marital	1
19.	monarchy	democracy	5
20.	seatback	backseat	8
21.	warfare	farewell	6
22.	smoking	hospital	7
23.	ape	ea	2

Некоторые пары заслуживают дополнительного комментария. Поскольку менять местами можно только **соседние** символы (это следует из пары *peat–tape* в условии: расстояние 4, а не 2), расстояние между *eel* и *lee* — 2, а не 1.

Поскольку все операции проводятся над одиночными символами, а не над произвольными последовательностями символов (это опять же видно из пары *peat–tape*), расстояние между *seatback* и *backseat* — не 1 (как интуитивно кажется человеку), и не 4, как нередко отвечают решатели, а 8.

Задание 3

Очевидно, расстояние Дамерау–Левенштейна не может превышать длину более длинного слова (заменой всех символов из него заведомо можно получить всё, что угодно), таким образом максимальное расстояние — m . Минимальное же расстояние — $m - n$ (число символов, которое необходимо добавить к короткому слову, чтобы получить длинное). Часто дается ответ 1, однако он не годится, так как в условии требуется выразить ответ через m и n .