



HAL
open science

A Statistical Approach to Persian Light Verb Constructions

Kim Gerdes, Pollet Samvelian

► **To cite this version:**

Kim Gerdes, Pollet Samvelian. A Statistical Approach to Persian Light Verb Constructions. 2008.
halshs-00668547

HAL Id: halshs-00668547

<https://shs.hal.science/halshs-00668547>

Preprint submitted on 9 Feb 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Statistical Approach to Persian Light Verb Constructions

Kim Gerdes¹, Pollet Samvelian²
Sorbonne Nouvelle

Abstract

This article presents the linguistic bases of Persian light verb constructions and shows the corpus based construction of lists of collocates for some common Persian verbs. The proposed methods of corpus construction are language independent and the good results on a relatively small corpus of 20 million words confirms the power of association measures based on the hypergeometric distribution. The resulting lists show a graduation of lexicalization and the semantic homogeneity of some light verb subcategorization schemes which could be the reason for their wide usage.

Keywords: Corpus, Persian, Light Verb Constructions, Statistics

1. Introduction

Persian light verb constructions have been extensively discussed in the literature, because of their abundance, productivity and particularities, but until recently all quantitative measures have been hampered by the lack of sufficiently large and sufficiently annotated corpora for this language. First, this paper shows why these constructions are particularly interesting for the description of Persian in particular, and for the study of collocation in general. Then we briefly describe the different steps taken to access Persian multiword expressions, using the GrosMoteur software: the construction of the corpus and of the conjugation tables and the choice and application of the association measure. Before concluding, we show and comment some of the abundant results and describe possible usage.

2. Persian Light Verb Constructions

Persian is well known for having very few simple verbs and for making a great use of multiword predicates, referred to as Complex Predicates (CPs) or Light Verb Constructions (LVCs), composed of a verb and a preverbal element. The number of simple verbs can be estimated around 200, but many of them do not display a high frequency of usage and are in competition with an LVC. Furthermore, despite the fact that morphological derivation is still available in Persian for creating new verbs, the verbal lexicon is mainly enriched by LVC creations and not by simple verb formation. This high degree of productivity entails that LVCs are not necessarily lexicalized and can be freely created on the basis of some semantic

¹ ILPGA, LPP (CNRS), Signes (Inria), kim.gerdes@univ-paris3.fr

² ILPGA, Mondes iranien et indien (CNRS), pollet.samvelian@univ-paris3.fr

and syntactic analogy with already lexicalized LVC patterns. Consequently, although the LVC phenomenon is not specific to Persian, the scope, the frequency and the productivity of these constructions in Persian raise some specific problems not necessarily encountered in other languages. This is the reason why Persian LVCs have received a great deal of attention during the last fifteen years in syntactic literature (Karimi-Doostan 1997, Samvelian 2001, Megerdooomian 2002, among others). However the main topic addressed in those studies concerns theoretical and representational issues. Thus, despite lexicographers' attempts to enrich Persian dictionaries with LVC entries (see for instance Sâdeqi 2000), a comprehensive description and semantic classification of Persian LVCs is still to be done and this task cannot be accomplished without a large scale corpora investigation.

This paper will focus on Noun-Verb combinations. Here is a non-exhaustive list of verbs participating in LVC formation: *âmadan* 'to come', *andâxtan* 'to throw', *âvardan* 'to bring', *bardâştan* 'to take', *bastan* 'to attach', 'to close', *bordan* 'to carry', *dâdan* 'to give', *dâştan* 'to have', *gereftan* 'to take', 'to receive', *gozâştan* 'to put', *kardan* 'to do', 'to make', *kešidan* 'to pull', *oftâdan* 'to fall', *raftan* 'to go', *šodan* 'to become', *xordan* 'to collide', *yâftan* 'to find', *zadan* 'to hit', *resândan* 'to bring', 'to make arrive', *residan* 'to arrive'. Some of these verbs, e.g. *kardan* 'to make', 'to do' or *šodan* 'to become', are prototypical light verbs (LV), in that they have a very poor semantic content and are always used as LVs. Many others, however, have also non-light uses. This is the case of *zadan* 'to hit, to fix, to apply':

- | | | | | |
|-----|---|-----------------------|----------|----------|
| (1) | N1 | N2 | Verb | |
| | maryam | omid-râ | zad | |
| | Maryam | Omid-DDO ³ | hit.PAST | |
| | 'Maryam hit Omid.' | | | |
| (2) | N1 | N2 | Prep-N | Verb |
| | maryam | tâblo-râ | be divâr | zad |
| | Maryam | picture-DDO | to wall | fix.PAST |
| | 'Maryam fixed the picture on the wall.' | | | |

As an LV, *zadan* occurs in various LVC constructions. In some cases, the valency pattern is identical to those already observed in the above examples, as in (3), but there are also patterns specific to *zadan* in LVCs, as in (4):

- | | | | | |
|-----|--|---------|--------------------------|----------|
| (3) | maryam | be omid | sili / labxand / telefon | zad |
| | Maryam | to Omid | slap / smile / telephone | hit.PAST |
| | 'Maryam slapped Omid.'; 'Maryam smiled to Omid.'; 'Maryam called Omid' | | | |
| (4) | tamâšâgar-ân | kaf | zad-and | |
| | spectators-PL | palm | hit.PAST-3.PL | |
| | 'The spectators applauded.' | | | |

Given the fact that it is generally assumed that the noun-verb combination in an LVC is idiomatic, it seems impossible at first sight to predict: first, which nouns combine with *zadan* and second, which is the appropriate construction when combining a given noun with *zadan*? Indeed, there are some examples like *labxand zadan* which satisfy different criteria suggested to identify LVCs and where the noun-verb combination is idiomatic: a) the noun phrase headed by the predicative and the LVC construction are synonymous; b) the subcategorization frame of the LVC is inherited from the noun; c) the first argument of the predicative noun is coreferent with the subject of the LVC. However, a more systematic investigation of the data reveals the two following points:

³ The enclitic *-râ*, which attaches to the right edge of noun phrases, marks the definite and/or specific direct objects in Persian.

1) In many cases there seems to be regularities in noun-verb combinations, in the sense that nouns with certain semantic and valence properties preferentially combine with a specific verb. Here are a few examples with *zadan*:

- a) «Scream nouns»: *faryād, jār, jiq, arbade*
- b) «Kick nouns»: *čak, kešide, kotak, lagad, sili* (kick, slap, beat)
- c) «Prejudice nouns»: *latme, sadame, zarbe, xesârat* (loss, suffering, damage)
- d) «Instrument nouns»: *bros, jâru, mesvâk, otu, šâne* (brush, broom, toothbrush, iron)

This provides evidence in favour of a semi-compositional approach for some Persian LVCs.

2) The nouns occurring in Persian LVCs are not necessarily predicative nouns and many ‘concrete’ nouns participate in LVC formation. This entails not only that Persian LVCs are much more various than LVCs *stricto sensu*, but also that there is no clear-cut distinction between LVCs and ordinary transitive and intransitive constructions. Compare:

- (5) maryam in vâks-râ be kafš-hâ-yaš zad
 Maryam this polish-DDO to show-PL-3.SG apply.PAST
 ‘Maryam applied this polish to her shows.’
- (6) maryam kafš-hâ-yaš-râ vâks zad
 Maryam show-PL-3.SG-DDO polish hit.PAST
 ‘Maryam polished her shows.’

These facts suggest that there is a continuum ranging from lexicalized LVCs to non-LVCs and that in intermediate cases, the ‘light’ verb is in fact a “semi” LV.

3. Corpus

As for all corpus work on multiword expressions, corpus based lexicon construction and quantitative measures on LVs are hampered by the lack of sufficiently large, and sufficiently annotated corpora, and the situation is particularly severe for Persian as virtually no free resources of considerable size are available with the notable exception of the Hamshahri Corpus (Darrudi et al. 2004). So why not work directly on the web with a search engine?

1. The main reason is that Persian is a language with relatively free word order and search engines basically offer research of cooccurrence of terms in a page, or adjacent strings. None of these choices can give any interesting results for a free word order language.
2. Moreover, Persian is an inflectional language which has to be accounted for in our statistical measures: it is possible to search individually for each of the nearly 60 forms per verb and the plural, possessive, and *Ezafe* forms of nouns, but to get relevant measures, one needs to cross all possibilities, unfeasible for any list of verbs.
3. Uncontrolled web research is generally error-prone but the problem is particularly important in Persian as the writing is less normative than many European languages, in particular concerning word spacing.

3.1 Corpus Construction

The present work is done with a specialized web crawler for linguistic corpus research, named «Grosporteur». The goal of Grosporteur is to give access to corpora without prior linguistic data and moreover, providing interfaces accessible to the common linguist without special computer training. Grosporteur is not specialized for the Persian language. So the same procedure can be applied to any language with sufficient Web resources. Grosporteur, mainly written in Python and Qt, allows for

- Crawling a specific site (or crawling freely with trainable language testing, which has proven useful for other languages and corpus needs)

All these steps can be controlled from Grosmoteur's graphical user interface. Grosmoteur also automatically creates an optimized word index giving direct access to all possible occurrences for each of the occurring words, which makes corpus access very fast in spite of its size.

4. Cumulative Hypergeometric Distribution for discovering collocations

We can now compute the collocates of certain verbs, chosen for their frequent use as support verbs, see the table of results on the last page. For a given infinitive and a window extension (in items to its left and right), Grosmoteur can find all occurrences of any of the verb forms, and the system extracts from the database all occurrences of all the forms with their corresponding windows. These joint windows will be our subcorpus. Our goal will be to compute the specificity of this subcorpus compared to the whole corpus.

Given a word with all its surrounding windows (the places where we expect the collocate), which are the words that occur more often than we would expect if the words were distributed evenly? The purely statistical answer is the cumulative hypergeometric distribution (Fisher's exact test). It avoids all expressive use of stop words or other ad-hoc measures, because it takes into account the total frequency of each word, the frequency of the word in the given windows and the size of the window. So if we stumble upon functional words in our result list, their presence is not an unwanted side effect but has to be explained syntactically, e.g. by a type of construction of the base verb. As Pedersen 1996 remarks, the computation of the hypergeometric distribution is prohibitively heavy and, at the time, was only applicable to small corpora. He shows that the results (at least for the dependent bigrams he's working on) are more reliable on sparse phenomenon like collocations. For our implementation, we developed a very effective computation of the cumulative hypergeometric distribution, with configurable thresholds.⁶

For each word in the subcorpus we apply the Fisher test, i.e. we compute the word's frequency in the subcorpus k and in the whole corpus m , and we compute the mode of the hypergeometric distribution (i.e. of the function giving the probability that the word appeared x times in this subcorpus). If k is bigger than the mode, we compute the sum H of the probabilities of the word appearing k or more times, if k is smaller than the mode, we compute the sum H of the probabilities of the word appearing k or less times. To distinguish probable from improbable events we return a negative result in the second case. Very small probabilities show "abnormal" correlations, positive values mean that the word appeared more than expected, negative values mean that the word appeared less than one might expect from the overall frequency of the word. Our tables show inverted "specificity" values ($1/H$) (cf. Lebart & Salem 1998) to get smooth results from high positive values to low negative values.

5. Results

As Persian is a verb final language, we chose a window of size 10 preceding the verb. Subsequent work will attempt to investigate the effects of variations of the window size. The table of results shows the first 70 most specific words with *andâxtan* 'to throw', *gereftan* 'to catch', *zadan* 'to hit' and *xordan* 'to collide' respectively. For each of these verbs, the majority of the most specific words are composed of nominal elements forming an LVC with the verb. However, some grammatical morphemes such as the definite object marker *râ* and a

⁶ Contrary to the rest of Grosmoteur, this function had to be written in C for performance reasons. We specialized a function from the GNU Scientific Library to our needs (<http://www.gnu.org/software/gsl/>).

few prepositions are also very high ranked. In this case, they provide relevant information about the syntactic construction of the LVC. The object marker indicates that the LVC construction is transitive. A preposition with a high specificity indicates either that the LVC construction contains a prepositional argument or that the nominal element of the LVC is itself introduced by a preposition. A few other high ranked words are semantically related to the most specific LVC constructions. For example, *šekast xordan* 'to be defeated' causes the high specificity of one of the construction's typical arguments: *došmanân* 'enemy'. For the sake of space, we comment only the most significant results for *zadan* and for *xordan*.

Comments on *zadan* 'to hit': The great majority of the words in the provided table are nominal collocates of *zadan* the few exceptions include prepositions used in the grammatical constructions of *zadan*. The following 500 entries, not shown here for lack of space, contain similarly good results. Then noisy results start to flood in. One of the highest ranked nouns is *dast* 'hand'. The combination of *dast* and *zadan* gives rise to two different LVCs, meaning 'to touch' and 'to accomplish':

- (7) a. maryam be miz dast zad b. maryam be xodkoši dast zad
 Maryam to table hand hit.past maryam to suicide hand hit.PAST
 'Maryam touched the table.' 'Maryam committed suicide.'

The first example is a simple LVC of *dast* and *zadan* and *be miz* is the argument of the construction. The second example shows the interesting compound LV formation: the combination of *dast* and *zadan* functions as a LV on its own, taking as a collocate a specific class of predicative nouns (e.g. *tazâhorât* 'demonstration'). The fact that *dast* is high ranked is mainly due to its use as a compound LV. Indeed, among the other high ranked nouns of the list, some occur only as an argument of *dast zadan*. To obtain the complete list of this specific class, we constrained our window to cases containing both *dast* and *zadan*. The specificity computation of this subwindow provided an extensive list of this semantic class. Here are the first 20 entries: *tazâhorât* 'demonstration', *be* 'to', *e'tesâb* 'strike', *tahasson* 'sit-in', *'aleyh* 'against', *gostarde* 'ample', *e'terâz* 'protestation', *râhpeymâyi* 'march', *jenâyat* 'crime', *eqdâm* 'action', *koštâr* 'killing', *kešwersân* 'countries', *xowdkoši* 'suicide', *čonin* 'such', *talâfi* 'revenge', *amaliât* 'actions', *xiâbân* 'street', *mosku* 'Moscow', *bozrog* 'big', *qatl* 'assassination', *mânovr* 'manoeuvre', *ahmaqâne* 'stupid', *šahr* 'city', *šuresš* 'rebellion'...

We see that most words belong to semantic class of «public or negative actions», the other being part of the grammatical construction or semantically linked (crime, demonstration → street, city ...) or are themselves collocates of the nouns (demonstration → big). The results are interesting for two reasons. First, they enrich the list of Persian LVs with compound LVs, which have never been investigated. Furthermore, they show that although the compound LV *dast zadan* can be translated as 'to accomplish', it obviously does not combine with all nouns designating an action, but only with a subset of them. Further investigations on other corpora are necessary in order to get a clear idea about the class of nouns combining with *dast zadan*, however, at this stage of investigation, these nouns roughly fall into three semantic classes:

- Nouns realizing the abstract concept of 'action': *amal* 'acte', *eqdâm* 'action'...
- Nouns designating public protest or rebellions: *tazâhorât* 'demonstration', *šuresš* 'riot'...
- Nouns designating crimes: *jenâyat* 'murder', *koštâr* 'massacre', *qatl-e âm* 'genocide'...

These facts constitute new evidence in favour of a semi-compositional approach to LVCs: speakers do not need to learn all combinations of nouns and LVs extensively. Instead, once they know some cases of well established examples of usage with a more or less regular form-meaning pairing they extend it to new cases judged similar to them.

Comments on *xordan* 'to collide': There are two homonymous verbs *xordan* in Persian, meaning 'to collide' or 'to eat'. Consequently, the specificity list contains both nouns forming an LVC with *xordan* and nouns designating comestible things. One of the highest ranked nouns is *šekast* 'defeat', which occurs in *šekast xordan* 'to suffer a defeat'. This is probably a

side-effect of our corpus: a political newspaper devoting several articles to different wars and conflicts in the Middle East. After a list of about a dozen predictive nouns, the most specific food noun is *qza* 'dish', which is a borderline case between free combinations (like 'eat pizza') and LVCs like *šekast xordan* 'to be defeated'.

6. Conclusion

We have shown how to obtain collocation lists only from large Web corpora with only minimal manual intervention even without prior linguistic resources. The corpus we collected is sufficiently large to provide extensive lists for some commonly used Persian verbs with the associative measure that we applied.

Work is in progress in various directions: 1. We are currently working on completing the morphological dictionaries of Persian, which will be provided as free resources. 2. We will have to diversify our corpus, maybe on blogs or internet forums in order to access more colloquial collocations. 3. We have already started to do reverse computations: Specificity computation of the windows for a given noun provides us with less used LV.

Our goal is the establishment of the productive schemes of Persian LVCs. We expect to contribute to a theoretical account of the gradual lexicalization of Persian LVCs, or more generally to the understanding of the widespread phenomena of collocations. A long term goal will be the elaboration of a linguistically well defined collocation dictionary of Persian in the style of Mel'cuk's Explicative and Combinatory Dictionary.

References

- CHURCH K. W. and HANKS P. (1990), «Word Association Norms, Mutual Information, and Lexicography», in *Computational Linguistics*, n°16-1.
- DARRUDI E., HEJAZI M. R. and OROUMCHIAN F. (2004), «Assessment of a Modern Farsi Corpus». *WITID2004*.
- FANO R. (1961), *Transmission of Information*, MIT Press, Cambridge, Mass.
- FISHER R. A (1970), *Statistical Methods for Research Research Workers*. Oliver & Boyd, Edinburgh, 14th edition.
- GOLDSMITH J. (2001), «Unsupervised learning of the morphology of a natural language», *Computational Linguistics* 27(2): 153-198.
- KARIMI-DOOSTAN M. (1997). *Light Verb Constructions in Persian*, PhD, University of Essex.
- LEBART L., SALEM A. and BERRY L. (1998). *Exploring Textual Data*, Kluwer.
- MEL'CUK I., ARBATCHEWSKY-JUMARIE N., IORDANSKAJA L., MANTHA S., POLGUERE A. (1984-1999), *Dictionnaire explicatif et combinatoire du français contemporain*. Recherches lexico-sémantiques, Montréal, Les Presses de l'Université de Montréal
- MEGERDOOMIAN K. (2002), *Beyond Words and Phrases: A Unified Theory of Predicate Composition*. PhD, University of Southern California.
- PEDERSEN T. (1996), «Fishing for exactness», in *Proceedings of the South-Central SAS Users Group Conference*, Austin, TX. www.d.umn.edu/~tpederse/Pubs/scsug96.pdf
- SAMVELIAN P. (2001), «Le statut syntaxique des objets nus en persan», *Bulletin de la Société de Linguistique de Paris*, n° 96-1 : 349-388.

SADEQI A. A. (2000), *Farhang-e fâris-e âmiyâne* (The Dictionary of Colloquial Persian). Niloufar Publications, Tehran.

Table of results

Pollet Samvelian 9/2/12 21:39
Mis en forme: Espace Avant : 6 pt

andâxtan 'to throw'		gereftan 'to catch'		zadan 'to hit'		xordan 'to collide'	
Spec	Word	Spec	Word	Spec	Word	Spec	Word
infinite	Rah	infinite	tsmim	infinite	xwd	infinite	skst
5E+259	Ra	infinite	tmas	6,14E+67	ra	infinite	čsm
7E+256	Bh	infinite	swrt	3,39E+35	dst	infinite	tlwtlw
2E+157	Xtr	infinite	qrar	1,53E+29	bh	1,34E+288	frib
4E+151	t'wiq	infinite	nzr	9,83E+20	xwab	2,93E+211	rqm
3E+103	Ngahi	infinite	nadidh	1,83E+20	rqm	2,19E+208	grh
3E+96	Grdn	infinite	mwrđ	6,58E+18	friad	4,32E+160	klid
6E+74	Mxatrđ	infinite	jlwi	6,26E+16	mhk	1,50E+156	w
3E+66	t'xir	infinite	drnzr	7,52E+13	hrf	1,42E+119	bh
6E+51	Lnrđ	infinite	dr	4,12E+13	w	5,86E+113	qsm
2E+49	Srš	infinite	badrnzr	1,98E+12	grh	2,41E+102	drđ
2E+46	Dam	infinite	amzaišan	5,6978E+11	damn	3,55E+89	gwl
1E+43	Brah	5,72E+297	afzaid	1,1162E+10	a'tsab	1,84E+83	qza
2E+42	Whšt	1,55E+292	axtiar	3599356765	astin	1,12E+82	tkan
9E+41	Čng	3,59E+266	ra	61100438,8	atš	4,27E+74	zrbh
6E+35	W	2,10E+223	br'hdh	20260470	brhm	1,80E+69	zd
4E+31	Saih	1,37E+218	tht	10981834,2	zrbh	1,25E+67	piwnd
2E+30	Ja	2,02E+154	fra	4170865,93	ja	2,10E+65	brhm
1E+23	Ab	2,17E+145	m'rz	4088629,32	sri	4,91E+63	zxm
7E+20	Pain	8,52E+136	piši	2252084,09	čhrh	5,38E+60	hm
7E+20	Jan	3,98E+130	'hdh	1703107,84	gl	1,07E+60	gl
1E+19	Taxir	3,26E+125	kar	1140782,56	srnwšt	2,64E+45	ab
1E+18	Swzn	9,96E+121	grwgan	561065,97	hirt	1,75E+40	ktk
1E+17	lad	5,41E+118	faslh	464976,928	piwnd	6,88E+39	hsrt
3E+16	mnař	1,63E+117	t'lq	245984,804	mhbwbđ	3,38E+38	nan
2E+14	'qb	6,92E+110	nšt	60115,209	qid	5,70E+36	ltmh
1E+14	Lrzh	6,59E+86	asart	15966,4507	ab	1,13E+36	dšmnan
9E+13	'lih	1,95E+85	dstwr	10038,4677	srbaz	9,80E+33	zmin
5E+13	Čřih	8,73E+77	jdi	8812,09557	br	6,32E+33	afsws
4E+13	Pñjh	3,89E+77	tsmimi	7276,51463	sazđ	9,05E+32	xwn
1E+13	Tnin	3,27E+73	brsi	4882,00645	zizlh	2,69E+32	mqađl
5E+12	Zndan	3,40E+71	hđf	4755,5065	framwš	5,65E+31	bhm
2E+12	Daxl	26537856,4	bkar	4358,91496	sprđh	3,31E+30	sili
1E+12	Jng	21782277	rdh	3930,00065	xanh	1,98E+30	mhk
9E+11	Qbqb	20444465,9	rwi	3871,11373	hfz	1,40E+30	t'sf
6E+11	Jlw	19812572,6	tm'	2628,83721	mhr	8,47E+28	xak
4E+11	Birwn	15971227,2	sraft	2572,11152	qlm	4,98E+28	qbth
3E+11	Drđsr	15918592,2	dwr	1670,76627	tzahrđ	3,68E+28	swgnd
2E+11	Nzri	15361670,7	aqwš	1501,57731	twanaii	2,38E+25	qsh
1E+11	Jnjali	11720420,3	čahi	1325,67147	mnař	1,09E+25	srnwšt
5,6265E+10	Tkapw	11273619,5	badi	1322,75903	snd	5,21E+24	bdrđ
4,0278E+10	Twp	9756447,09	tbligati	1152,82085	brai	2,66E+22	ntijh
2,6024E+10	Rwdxanh	7073116,01	aw	883,396108	sr	7,90E+20	glhai
1,8325E+10	Dhan	5964328,85	čah	741,051801	bzrgi	2,51E+20	čwb
1,0793E+10	Jsdš	5481239,84	krna	737,484023	jaigah	8,45E+19	astart
7524844172	Sanh	5419321,52	sršan	622,189263	dwr	4,11E+19	trkš
6595098377	Andam	3778400,05	hras	534,574011	gstrđh	1,81E+19	tim
5731891108	Iarw	2955678,15	slanti	475,394426	bala	1,29E+19	az
4092357198	Dwš	1990392,02	aštbađ	342,010514	bhran	6,59E+17	sbhanđ
2490665981	Tqsir	1870854,77	brgrđn	257,714115	saz	2,99E+17	liz
1967075134	Sndwq	1864685,17	fkr	248,816392	sahb	9,27E+16	xt
1450832614	Grđnš	1573105,51	dstš	235,519022	xwastđ	2,77E+16	brgšt
1392405218	Wlwlđ	957212,23	jar	229,761253	bimđ	2,54E+16	trk
1235788952	Psrk	720810,963	mxmšđ	225,72005	hmwarđ	8,30E+15	dl
1220608635	Črťkh	646896,549	bsat	210,404618	jai	2,16E+14	dannd
969943866	Rwani	314142,215	dria	206,158025	hqwq	6,15E+13	wwl
648582030	Tzahrđ	308930,492	dwid	107,835242	asib	6,13E+13	brabr
647318980	Zhmt	288452,011	hirwšima	93,7371696	hsab	1,82E+13	sxti
417725451	Sr	280125,105	ws'i	85,9155486	mlt	1,54E+13	nhar
413329285	pa'in	202366,785	jnjali	84,9936369	salđ	1,25E+13	zrbt
236999204	Jrian	187440,188	ksif	84,2853335	thrim	1,12E+13	bčsm
84416671	Axtlař	160693,141	faslh	76,9412072	'ml	9,80E+12	hrif
77876871,2	tw'tđ	157986,747	šah	76,6266683	zmin	9,72E+12	lqmđ