Towards a Description of Urdu Multiword Items for English Learners: A Lexicological Perspective

[1]*Abrar Hussain Qureshi*
[2]*Shamim Akhter*
[3]*Akbar Sajid*

Abstract

There is a considerable shift in foreign language learning from single word lexical items to multiword items, also known as phraseology. As meaning is present in larger stretch of language, the significance of these multi-word items of language has increased many times. To comprehend a phraseological unit is attached with the probability of our better understanding about the world diachronically. Multiword items include idioms, collocations, set phrases, proverbs, etc. Anyhow, handling of collocations and idioms for the learners of a foreign language is very challenging. The undertaken study is a systematic attempt in this regard to describe Urdu language on the level of lexical collocations and idioms. The corpus of Urdu language "urTenTen: Corpus of Urdu Web" has been used for the description of Urdu lexical collocation. To process the corpus data, "Sketch Engine" has been used. The retrieval of Urdu lexical collocations and idioms are supposed to be a rich resource for Urdu lexicologists, lexicographers, grammarians, language learners and teachers, etc.

Keywords: Multi-word items, description, corpus, collocations, idioms
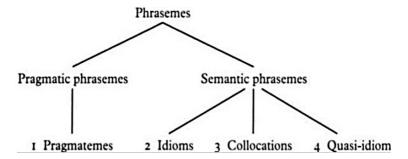
1. Introduction

Generally, the concept of multi-word items reminds the idea of collocations and idioms though their range is a bit larger and even cover the concept of maxims, compounds, set phrases, etc. as well. According to Hussain and Iqbal (2009) multi-word items are the lexical items whose meaning is often conceived more than the meaning of their constituents. Both attach special significance to collocation and idioms as the strongest form of multi-word items. The phenomenon of multi-word items occurs when some words frequently and regularly combines with some other word or with some grammatical construction (Romer, 2009).

---

[1] Assistant Professor of English University of Sahiwal
[2] PhD scholar, School of Languages, Civilization and Philosophy University Utara Malaysia
[3] Assistant Professor of English NUML Multan

Masini (2005) asserts that multi-word items are larger than a single word that sometimes has compositional and idiomatic meanings. He continues that theoretically they cover a wider range from free collocations to fixed expressions. Fixed expressions generally have opaque meanings and are stored in the mental lexicon of the native speakers. They are non-compositional due to the fact they cannot be modified as far as their internal structure is concerned. According to Mel'čuk. (1998) the term "phraseology" is used to refer to the phenomenon of multi-word items. Though, relatively, it is a new concept and a new branch of linguistics yet, it covers the possibly most of the concepts that multi-word items carry. Mel'čuk ranges his discussion of phraseology by and large between idioms and collocation. He discusses phraseology on semantic basis with reference to idioms, collocations and quasi-idioms. He excludes structural and grammatical consideration in the study of phraseology or multiword items.

Typology of phraseology by Mel'čuk (1998)

As it has already been discussed that there is a wide range of multi-word items, but they do not meet the criteria of one comprehensive definition. The following list gives us a general overview of the concept of multi-word items and it should not be taken as the classification, for example, multiword items range from routine formula, collocations, frozen forms, complex nominal, bio nominal expressions, phrasal verbs, and idioms to proverbs. This list is not applicable in all cases. It confuses criteria of semantic, syntactic,

23

pragmatic explanations of words (Amosova, 2013). Nevertheless, it provides a range of multi-word items.

With the advent of computers and large corpora, the studies regarding multi-word units have got new dimensions and are based on large frequency data rather than the prevalent orthodox linguistic criteria of defining multi-word items (Fernando, 1996). Similarly, the concept of multi-word items has been explained by various lexicologists, more or less in the same spirit but in the following discussion, the term will be used with the phenomenon of semantic compositionality and syntactic fixedness.

Moreover, it is not possible to deal with all the multi-word items in one article as it will out space the focus of the undertaken research. Consequently, two levels of the Urdu multi-word items have been chosen for analysis that is collocation and idiom of the Urdu language.

1.1 Multiword items and Lexicology
The research regarding multi-word items has got special significance in the field of lexicology and second language learning. As corpus linguistics has shaped into a highly a scientific linguistic inquiry, it has also innovated the art and craft of comparative lexicology (Fernando, 1996). Rather, corpus linguistic has revolutionized the theory and practices of lexicology. Traditional lexicology focused on single word lexical item but with the recent advancement in linguistic studies across the globe, there is a considerable shift from single word lexical items to multi-word lexical items. Consequently, multi-word items in the Urdu language have got a significant place in comparative lexicology (Akhtar, 2003).

1.2 Description of Urdu Idioms
It goes without saying that handling of idioms poses a great problem not only to the language learners but also to the lexicologists (Glucksberg, 2001). Idioms are an important level of multi-word items and can be classified as a complex multi-word items as to comprehend them is a challenging task semantically and structurally (Fellbaum et al., 2006:239) but on the basis of semantic unity, idioms are considered lexical items and are defined as a multi-word unit, emerged by frequent use and develop a meaning that cannot be deduced from its constituents.

According to Sinclair, (1997) structural description of idioms and collocations include both grammar and lexis and can be placed on the same continuum, though idioms have strict semantic and structural fixedness and are at the one side of the continuum while collocations have the semantic and structural looseness and are placed on the other side of the continuum.

Lexical Level (Collocation and Idioms)

	Syntagmatic	Paradigmatic
Grammar (syntax)	Structure (e.g. SVO, dhq, SPOCA)	System (e.g. *pronoun system; active vs. passive*)
Lexis (vocabulary)	Collocation (e.g. *rancid butter, addled eggs, stale bread*)	Sets (e.g. *lexical field of vehicles, flowers, etc.*)

Sinclair's description of Lexical level (1997)

Most idioms are learned and re-used as single lexical item, yet they are not single words. The following Urdu idioms comprise three words but they are single lexical items. Idioms in the Urdu language may be modified in respect of gender and number.

ا نگا ر و ں پہ لیٹنا   To be on pins and needles

چرا غ گل ہو نا   End of one's life

1.3 Semantic Criteria

Linguists have devised many methods to describe idioms. The most authentic way to describe idioms is undoubtedly their meanings. According to Bailey (1932) Urdu idioms may be described in terms of their structural

and internal complexities. Meaning of some Urdu idioms cannot be deduced from its parts, are called complex idioms, for example گل کھلا نا that literally means to blossom a flower but the real meaning of the idiom is to create problem. Some idioms are semi-complex in the Urdu language. They are less unintelligible and some meaning may be worked out from the constituents of the idiom for example انگا ر و ں پہ لیٹنا. The literal meaning of the idiom is to walk on the burning coals and the figurative meaning is to be in serious problem. So, the meaning can be worked out in semi-complex idioms. The third type is the transparent idioms in the Urdu language, where meaning of the idiom can be easily guessed by its constituents for example. د ل د کھا نا The literal and the figurative meaning of the idiom is the same.

1.4 Syntactical Criteria

Second parameter to describe idioms in Urdu language is the syntactic dimension. However, syntactic dimension of Urdu idioms can help the Urdu lexicologist. The syntactic patterns of Urdu idioms are very diverse but a few syntactical patterns occur frequently, for example, noun verb combination is very common in Urdu idioms as آ نکھیں د کھا نا. Another combination is noun preposition noun for example آ ستین کاسانپ. Another frequent syntactical combination is adjective noun for example, کا لا چو ر.

1.5 Grammatical Criteria

Third dimension of description of Urdu idioms is grammatical classification. Idioms in the Urdu language may be described in terms of parts of speech. Some Urdu idioms can be grammatically labeled as verbal for example چا ندیبو نا (To get huge profit). Some idioms can be labeled as adverbial for example پلک جھپکتے ہی (In no time) while some Urdu idioms can be indexed as adjectival for example د ہو ا ں د ھار (emotional). Most idioms in the Urdu language are nominal in which the head is the noun for example کا لا چو ر (an unknown person)

The concept of key headword can prove very useful in indexing idioms in an Urdu-English bilingual dictionary. The key head words should be indexed alphabetically and each head word should show the cluster of idioms that are attached with that head word or with various parts of speech with which that head word of an idiom may occur (Amosova, 2013) for example:

چا ندیہو نا	To become silver	To get huge profit
عید کا چا ند ہو نا	To be the moon of	once in a blue moon

An Urdu English bilingual dictionary should accept a wide range of Urdu idiomatic expressions.

Urdu idioms are of many types but according to the current lexicological norms, they are indexed alphabetically rather than according to their semantic or syntactic group. Ilson (2002) asserts that while entering idioms in a dictionary, a semasiological approach that covers from name to notion, should be followed.

1.6 Description of Urdu Collocations

Collocations are yet another significant level of lexis in the Urdu Language. Habitual co-occurrence of two or more words is called "collocation" (Firth 1957). The structure of a collocation consists of a word that frequently prefers a specific lexical realization of the idea, the other word represents. Like other languages of the world, Urdu also allows many frequent habitual co-occurrence of words e.g. اچھا *accha* (good) can have freer combination with many nouns اچھا گھر *acchaghar* (good House), اچھا لڑ کا *acchalarka* (good boy), etc.

simple **9,625** (158.28 per million) اچھا

Collocations

	Word	Cooccurrences	Candidates	T-score	MI	LogDice		Word	Cooccurrences	Candidates	T-score	MI	LogDice
1	لگتا	388	7,654	19.64	8.32	9.52	11	برتاؤ	53	480	7.27	9.45	7.43
2	خاصا	171	779	13.07	10.44	9.07	12	سا	104	10,287	10.04	6.00	7.42
3	بہت	1,189	72,443	34.15	6.70	8.89	13	لگند	59	1,781	7.64	7.71	7.41
4	سلوک	137	3,140	11.66	8.11	8.46	14	لکھا	118	13,511	10.67	5.79	7.38
5	لگا	294	19,648	16.97	6.56	8.36	15	اقدام	67	3,940	8.11	6.75	7.34
6	اچھا	178	9,625	13.23	6.87	8.24	16	لگے	97	10,872	9.67	5.82	7.28
7	کتنا	114	3,304	10.63	7.77	8.17	17	بوتا	282	52,221	16.30	5.09	7.22
8	برا	116	3,696	10.72	7.63	8.16	18	مجھے	237	43,066	14.95	5.12	7.20
9	اتنا	132	11,281	11.33	6.21	7.69	19	طریقہ	76	8,082	8.57	5.89	7.14
10	کام	352	53,358	18.31	5.38	7.52	20	بمارا	84	10,464	8.98	5.66	7.10

A few words in the Urdu language allow very restricted company of words. For instance, the adjective با سی baasi (stale) makes a collocation like با سی ر *baasi roti* (stale bread) روٹی با سی سبز ی *baasisabzi* (stale vegetable) but does not make a collocation with some uncooked edibles like گھیو ں *gehu* (ghee) as با سی گھیو ں*baasigehu* making such structures unacceptable in the Urdu language. Rather, some words co-occur so frequently that they form a single entity syntactically and semantically. The node word **" نظر"** is restricted neither on semantic range nor or meaning but confined to one or two words only "e.g. " سر سر ی or . سطحی

1.7 Types of Collocations in Urdu

Collocations in the Urdu language may be described form the following perspectives. Firstly, some collocations in the Urdu language are unrestricted. This means that a node can occur with a wide range of lexical items (Romer, 2009). This involves core vocabulary of the Urdu language. For example, پا نی(water) can collocate with a vast range of living and non-living, concrete and abstract items: e.g. پا نی لگا نا (to water), پا نی پینا (to drink water), پا نی ڈ ا لنا to throw water), پا نی پلا نا etc. However, there are some hazy areas between free combinations and idioms/collocations. The following concordance of the word پا نیproves that it can occur unrestrictedly with a vast range of words.

simple پانی **23,431** (385.32 per million)

Collocations

Word	Cooccurrences	Candidates	T-score	MI	LogDice	Word	Cooccurrences	Candidates	T-score	MI	LogDice
1 پینے	824	3,111	28.66	9.42	9.99	11 پانی	463	23,431	21.10	5.68	8.34
2 سیلابی	691	2,472	26.25	9.50	9.77	12 اخراج	230	1,014	15.14	9.20	8.27
3 صاف	695	7,765	26.25	7.86	9.51	13 قلت	249	3,494	15.69	7.53	8.24
4 سطح	702	9,175	26.36	7.63	9.46	14 دریائی	237	3,325	15.31	7.53	8.18
5 بجلی	601	10,277	24.35	7.25	9.19	15 قلت	207	1,271	14.35	8.72	8.10
6 جھیل	305	2,084	17.42	8.57	8.61	16 ڈوب	195	1,587	13.92	8.32	8.00
7 بیراج	283	1,199	16.80	9.26	8.56	17 آمد	218	4,671	14.64	6.92	7.99
8 کیوسک	275	865	16.56	9.69	8.53	18 کمی	245	8,253	15.45	6.27	7.99
9 گرم	289	3,430	16.92	7.77	8.46	19 ٹیم	191	3,297	13.73	7.23	7.87
10 بہہ	258	1,734	16.02	8.59	8.39	20 فراہمی	178	2,429	13.27	7.57	7.82

Secondly are the collocations in the Urdu language that are semi-restricted. In these collocations, node allows comparatively determined options. For example, the possible combinations with the node آ سما نی (heavenly) are not large in number, e.g. آ سما نی کتب (heavenly books), آ سما نی ر حمتیں (heavenly blessings), آ سما نی آ فتیں (heavenly sufferings).

simple آسمانی **896** (14.73 per million)

Collocations

	Word	Cooccurrences	Candidates	T-score	MI	LogDice
1	کتابوں	79	4,152	8.88	10.33	9.00
2	آفات	26	793	5.10	11.12	8.98
3	،فرشتوں	12	22	3.46	15.18	8.74
4	ادیان	16	329	4.00	11.69	8.74
5	صحیفہ	14	281	3.74	11.72	8.61
6	صحیفے	12	123	3.46	12.69	8.59
7	صحیفوں	8	73	2.83	12.86	8.08
8	،کتابیں	7	61	2.65	12.93	7.90
9	بجلی	80	10,277	8.93	9.05	7.87
10	کتابیں	28	3,158	5.28	9.23	7.82

Thirdly are the collocations in the Urdu language that are extremely restricted. These do not allow free combinations with the node. Rather, only one or two combinations are possible e.g. چلچلا تی دھو پ (scorching sun), چکنی باتیں چو پڑ ی (flattering words), گورا ر نگ (*white color*).

simple چکنی چیڑی **31** (0.51 per million)

Collocations

	Word	Cooccurrences	Candidates	T-score	MI	LogDice
1	باتوں	23	5,626	4.80	12.97	7.06
2	باتیں	4	9,770	2.00	9.65	3.74
3	اپنی	3	95,738	1.70	5.94	0.04
4	کی	19	1,309,420	4.21	4.83	-1.07

simple چلچلا ئی **2** > shuffle **2** (0.03 per million) ✕

Details		Left context	KWIC	Right context
1 nooremadinah.net	م علیہ اللہ صلی رسول رت زیامیں می گر کی حرہ دن را سار اسامیں دھوپ **تی چلچلا** جو لوگ تے ہ وہ ہبا آنسو اور تے کر یادکو وسلم علیہ اللہ صلی اتھتی۔عشاق،
2 sulemansubhani.wordpress.c...	م علیہ اللہ صلی رسول رت زیامیں می گر کی حرہ دن را سار اسامیں دھوپ **تی چلچلا** جو لوگ تے ہ وہ ہبا آنسو اور تے کر یادکو وسلم علیہ اللہ صلی اتھتی۔عشاق،

Lastly are the collocations in the Urdu language that are inherent. For example, some activities are specific with the human as (ر و نا)(to weep) is purely a human phenomenon. These types of collocation make them absolutely inherent.

1.8 Lexical Collocations in Urdu

1.8.1 Noun-Noun

Like English, there are a few patterns of lexical collocations in the Urdu language. The first lexical pattern of Urdu collocations is Noun Noun collocation that is quite frequent.

simple صدر **39,626** (651.65 per million)

Collocations   CHANGE CRITERIA   BACK TO CONCORDANCE

Word	Cooccurrences	Candidates	T-score	MI	LogDice	Word	Cooccurrences	Candidates	T-score	MI	LogDice
1 زرداری	7,176	13,850	84.60	9.63	12.10	11 علی	2,277	36,090	47.23	6.60	9.94
2 آصف	4,596	9,748	67.70	9.50	11.57	12 مملکت	1,066	3,247	32.58	8.98	9.67
3 اوباما	3,621	5,750	60.11	9.92	11.35	13 بش	962	2,016	30.97	9.52	9.56
4 امریکی	3,841	29,584	61.66	7.64	10.83	14 نائب	866	2,875	29.36	8.85	9.38
5 باراک	2,297	2,879	47.89	10.26	10.79	15 احمدی	750	2,480	27.33	8.86	9.19
6 مشرف	2,777	12,064	52.55	8.46	10.78	16 نژاد	724	2,506	26.85	8.79	9.14
7 جارجی	2,030	6,065	44.97	9.00	10.51	17 گیا	2,348	98,440	47.13	5.19	9.12
8 ایوان	1,416	3,728	37.57	9.19	10.06	18 کرزئی	664	1,408	25.73	9.50	9.05
9 پرویز	1,484	7,612	38.39	8.22	10.01	19 جنرل	873	19,483	29.12	6.10	8.92
10 سابق	1,591	11,090	39.71	7.78	10.01	20 نے	9,374	600,351	92.78	4.58	8.91

The above data shows that Urdu noun "صدر (president)" makes free combination with other Urdu nouns. The fact is also noteworthy that nouns occur in lot of combinations of lexical collocations in most of the languages of the world. Some compounds may also be distinguished from noun noun combinations and there needs an extreme caution on the part of Urdu

learners to comprehend the difference between the two. For example, "صدرمملکت" (president of country) sounds more an Urdu compound owing to its comparative frozenness. According to Marchand (1969) a compound must be morphologically isolated from parallel syntactic group.

1.8.2 Noun-Verb
The second lexical pattern of Urdu collocation is Noun Verb that is also very common.

	پانی	
نوشیرواں گیا دوڑ گیا چھوڑ جنگل اپنا سب ہ متاع مال گیا موڑ میرا نوشیرواں میدانوں خواب حال ڈ	پانی	ھا ویکھیا مک وچ راہ پا نصرت ایپ بن دی وطن مڑ یا تمام اسباب تے لٹیا کیانی تخت بشام فتح
اتنا اوپر کے وسلم علیہ اللہ صلی آپ وقت اس لگے۔گیا دیا بٹھا میں لگن ایک کو وسلم علیہ اللہ صلی آپ ہونے ک	پانی	کچھ نے وسلم علیہ اللہ صلی آپ کو ڈال گیا ڈالا ۔ " بس بس" وسلم علیہ اللہ صلی آپ کہنے
میں لگن لیے میرے فرمایا نے وسلم علیہ اللہ صلی آپ ۔"بیں رہے کر کا انتظار کا آپ سب۔ اللہ رسول؛	پانی	پ لیکن چابا اٹھنا بعد کے اس اور فرمایا غل نے وسلم علیہ اللہ صلی آپ۔ کیا ہی ایسانے ہم۔ رکھو
میں کثرے سامنے کے وسلم علیہ اللہ صلی آپ کی۔ مسواک طرح اچھی نہایت نے وسلم علیہ اللہ ص	پانی	تے فرماتے اور تھے پونچھتے چہرہ کر ڈال ہاتھ دونوں میں پانی وسلم علیہ اللہ صلی آپ۔ تھا
وسلم علیہ اللہ صلی آپ۔ تھا پانی میں کثرے سامنے کے وسلم علیہ اللہ صلی آپ کی۔ مسواک طرح	پانی	یکے اللہ "اللہ الہ الا لا" نہیں جاتے فرماتے اور تھے جاتے پونچھتے چہرہ کر ڈال ہاتھ دونوں میں
دن کے بقت مجھیاں پاس کے ان جب لگے سے ان حد میں معاملہ کے بقت جب تھی پر کنارے	پانی	وہ لیے اس آزمایا طرح اس انہیں نے ہم تھیں آتی نہ تو بو نہ بقت دن جس اور لگیں آنے جن کے
سے اس نی قوم کی اس جب بھیجا حکم نے ہم کو موسیٰ اور تھیں جماعتیں بڑی بڑی جو اولادہ کی د	پانی	لن گھاٹ اپنی نے نکلے پر کینجھر میں چشمے بارہ سے اس تو مار یہ پتھر اس لاٹھی اپنی کہ مانگا
میٹھے کی کراچی خصوصا شہروں ملحقہ سے جہاں بیں واقع جھیلیں کئی میں ٹھہ۔ ضلع <s><s>ا	پانی	ٹیٹ معروف پر سطح قومی بالیجی اور کینجھر میں جھیلوں ان بیں۔ جاتی کی پوری ضروریات کی
بالٹی سو دو سے کنویں ; 8220# کہ بتایا یہ حل کا اس بعد کے جاتنے ماجرا سارا انے صاحب م	پانی	پانی بعد دنوں کچھ مگر گئی دی نکال پانی بالٹی سے دو ; 8220# گا جائے بو پاک کنواں نکالو
بالٹی سے دو ; 8220# گا جائے بو پاک کنواں نکالو پانی دی نکال	پانی	برا کیا یہ کہ گئے بو پریشان لوگ گئے کہ ماجرا کیا اس وہ تو بو شروع گی آنی بدو سے پانی
بعد دنوں کچھ مگر گئی دی نکال پانی بالٹی سے دو ; 8220# گا جائے بو پاک کنواں نکالو پانی،	پانی	ر ایک اور لئے کے اس وہ تو بڑھی ماجرا کیا یہ کہ گئے بو پریشان لوگ گئے کہ ماجرا کیا اس وہ
بالٹی سو چار سے کنویں ; 8220# کہا نے انہوں تو گئے پاس کے مولوی اور ایک اور لئے کے	پانی	نی مگر گئی دی نکال پانی بالٹی سو چار ; 8220# گا جائے بو ٹھیک سب ; 1548# دو نکال
ر حل کے اس وہ تو بڑھی پریشانی کی لوگوں ; 1748# گئی نہ بدو سے پانی مگر گئی دی نکال	پانی	ح حل کے اس وہ تو بڑھی پریشانی کی لوگوں ; 1748# گئی نہ بدو سے پانی مگر گئی دی نکال
مولوی تیسرے لئے کے حل کے اس وہ تو بڑھی پریشانی کی لوگوں ; 1748# گئی نہ بدو سے	پانی	مگر گئی دی نکال پانی بالٹی سو چار ; 8220# گا جائے بو ٹھیک سب ; 1548# دو نکال پ

Allerton (1984) asserts that in Noun verb collocations, noun carries the main information while verb is a part of the collocation structure. Urdu Nouns like other languages of the word cover the bulk of the Urdu language. It occurs with other parts of speech quite frequently, creating a rich source of Urdu collocations. The noun verb combination is as unrestricted as adjective noun combination in the Urdu language.

1.8.3 Adjective-Noun
Third combination of lexical collocation in the Urdu language is Adjective Noun. Urdu nouns are accompanied by typical adjectives.

Collocations

Word	Cooccurrences	Candidates	T-score	MI	LogDice
نبایت 1	79	7,014	8.83	7.26	7.82
انداز 2	106	12,787	10.20	6.82	7.65
انتہائی 3	69	9,974	8.22	6.56	7.29
ترین 4	81	13,745	8.89	6.33	7.19
گیت 5	25	1,220	4.98	8.13	7.17
بت 6	310	72,443	17.31	5.87	7.05
غزل 7	30	3,192	5.43	7.00	7.01
تفریحی 8	16	314	3.99	9.44	6.78
مناظر 9	19	1,338	4.34	7.60	6.75
لڑکی 10	29	5,061	5.32	6.29	6.64

Word	Cooccurrences	Candidates	T-score	MI	LogDice
اشعار 11	21	2,462	4.54	6.86	6.64
عمارت 12	23	3,208	4.75	6.61	6.62
گنبد 13	14	593	3.73	8.33	6.51
لباس 14	23	4,149	4.73	6.24	6.45
عورت 15	52	15,398	7.05	5.53	6.42
نظم 16	23	4,368	4.73	6.17	6.42
منظر 17	30	7,171	5.38	5.83	6.40
اتنی 18	35	10,643	5.78	5.49	6.25
لڑکیوں 19	15	2,097	3.83	6.61	6.23
لگ 20	26	7,286	4.99	5.61	6.18

The above combination is commonly found in the Urdu language due its structural looseness. The node "خوبصورت" is making free combinations with a lot of nouns like " مناظر, لڑکی ,عورت ,لباس " , etc.

1.8.4 Adverb-Adjective

The Fourth possible pattern of lexical combination in the Urdu language is Adverb Adjective. This pattern of lexical collocation is not as frequent as the rest of lexical patterns of Urdu collocations.

simple بہت **72,443** (1,191.32 per million)

Collocations

	Word	Cooccurrences	Candidates	T-score	MI	LogDice		Word	Cooccurrences	Candidates	T-score	MI	LogDice
1	سی	3,392	44,104	57.34	6.01	9.90	11	اچھی	846	9,169	28.71	6.28	8.41
2	زیادہ	3,605	52,040	59.01	5.86	9.89	12	کچھ	1,554	78,326	37.05	4.06	8.40
3	بڑا	2,041	17,910	44.71	6.58	9.53	13	شکریہ	796	6,769	27.93	6.63	8.36
4	کم	2,261	36,597	46.63	5.70	9.41	14	مشکل	836	10,801	28.47	6.02	8.36
5	بڑی	1,741	27,181	40.95	5.75	9.16	15	خوش	753	12,087	26.92	5.71	8.19
6	سارے	1,256	9,828	35.11	6.75	8.97	16	بڑے	803	23,195	27.36	4.86	8.10
7	بی	3,420	162,911	55.16	4.14	8.90	17	ام	763	19,351	26.79	5.05	8.09
8	جلد	1,119	12,325	33.01	6.25	8.76	18	ضروری	719	19,082	25.97	4.98	8.01
9	اچھا	1,058	9,625	32.17	6.53	8.72	19	بہت	1,116	72,443	30.82	3.69	7.98
10	سے	12,276	969,166	100.38	3.41	8.59	20	ساری	583	9,434	23.68	5.70	7.87

The above concordance shows the limited number of lexical combinations of adverb and adjective in the Urdu language. Some other lexical patterns of collocations in the Urdu language may be found but their frequency is very restricted as no general structure of lexical pattern can be determined.

2. Discussion

Structurally, Urdu is a different language from English. The significance of such descriptive studies becomes all the more important when the non-native learners of the Urdu language constantly make comparisons of the target language with their mother tongue. Realistically, such convergence of the target language and the mother tongue is not potential as the Urdu collocation "یسوع مسیح" and "Jesus Christ" have difference of meaning when used in their respective cultures. Conversely, these attempts of convergence by English learners may lead to unsuccessful handling of the Urdu language.

Predictably, Urdu collocations described in the above section, can be classified the most frequent as they have been retrieved from large corpus data. Without considerable awareness of collocations, as supported by Lewis (2004), it is difficult to handle overall language skills of Urdu. These are the collocations that shape the linguistic expressions of non-native learners, natural and fluent even though they lack grammatical strength sometimes. However, it is an established fact that Urdu as L1 and as L2, lexical

development differs significantly. The first difference in the case of Urdu as L2 is the poverty of input of the word combinations in terms of quality and quantity that make the bulk of Urdu lexicon.

The L2 learners of Urdu have to face lack of contextualized data that can only be presented in the form of lexical collocations. This situation puts foreign learners of Urdu language in an extremely difficult position to create syntactic, semantic and morphological harmony among word associations of Urdu language and they are at a loss to integrate words into acceptable combinations. Learners of Urdu as a foreign language usually focus on the single words items and disregard their association and context. They try to learn lexical collocations of Urdu as separate words rather than in chunks. Consequently, they often refer to their mother tongue to produce word combinations in the target language that results in unacceptable lexical choices.

Summing up the corpus data, it can be safely claimed that Urdu is a very rich language. Its lexicon structure is very diverse. Urdu language offers many word combinations of lexical collocations. The corpus data of Urdu lexical collocation can be very useful in the description of Urdu language at the level of lexis. Apart from helping the foreign language learners (Kunin, 2005), descriptive studies help the translators to translate books for the foreign learners with lexical sensitivity, for the grammarians who intend to publish books on lexis and in language teaching for native speakers as well as for the foreign learners of the language.

3. Concluding Remarks

The undertaken research has investigated multi-word lexical items in the Urdu language. The real motive behind this lexical study of the Urdu language is to examine the Urdu lexis from lexicological point of view. As, there is a dearth of lexical work in the Urdu language, it was the need of the hour to carry out this research to facilitate the teachers, learners, and the

lexicologists. However, a general lexical study may be useful in various fields of linguistic research (Haspelmath, 2002).

The present study aims to analyse the lexicology of Urdu in terms of multiword items. The usefulness of the present study lies in the fact that it can be employed in various fields like semantics, sociology of language, stylistics, compiling dictionaries, teaching of Urdu as a second language (TUSL) and preparing materials for different levels i.e. primary, secondary and tertiary levels. The undertaken research will help the lexicologists of the Urdu Language to understand the nature and the formation of the Urdu lexis. It will also be a source of inspiration for Urdu lexicographers to look at the nature of Urdu lexis in detail before compiling an Urdu monolingual or bilingual dictionary involving the Urdu language as the source or the target language.

In the end all the dimensions of Urdu idioms and collocations may not have been investigated. However, the research can be taken as a point of departure and motivates the Urdu lexicologists and lexicographers to further explore Urdu idioms and collocations as a rich linguistic resource of the Urdu language.

References

Akhtar, S. (2003). *Urdu ZubanKya Hay?* Lahore: Sang-e-Meel Publishers.

Alerton, D.J. (1984). Three or Four Levels of Co-occurrence Relations. *Lingua* 63:17-40.

Amosova, N.N. (2013). Fundamentals of English phraseology. Moscow. Librokom.

Bailey, G. (1932). *A History of Urdu Literature*. London.

Fellbaum, C. et al (2006). Corpus-based studies of German Idioms and light verbs. *International journal of lexicography* 19(4). 349-360

Fernando, C. (1996). *Idioms and Idiomaticity*. Oxford: OUP.

Firth, J.R. (1951). *A Synopsys of Linguistic theory*. Oxford: Blackwell.

Glucksberg, S. (2001). *Understanding Figurative Language: From Metaphors to Idioms*. Oxford: OUP.

Haspelmath, M. (2002). *Understanding Morphology*. London: Arnold OUP.

Hoey, M. (2005). *Lexical priming*. New York, NY: Routledge.

Hussain, A. and Iqbal, Z. (2009). A Corpus Based Study of Urdu English Lexicography' 2009. *ELF Annual Research Journal* Vol 11 Shah Abdul Latif

Kunin, A.V. (2005). *A course on modern English phraseology*. Dubna: Feniks.

Masini, CP. (2005). *The Indo-Aryan languages*. Cambridge University Press

MEL'ČUK, I. (1998). Collocations and Lexical Functions, in A. P. Cowie [ed.], *Phraseology*

Marchand, H. (1969). *The Categories and Types of Present-Day English Word-Formation*. 2nd revised edition. München: C. H. Beck.

Romer, U. (2009). *The inseparability of lexis and grammar. Annual Review of Cognitive Linguistics*.

Sinclair, J. M. (1997). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.